Emotion Perception from Face, Voice, and Touch: Comparisons and Convergence

Annett Schirmer[1] and **Ralph Adolphs**[2]

[1]National University of Singapore, Singapore

[2]California Institute of Technology, USA

Abstract

Historically, research on emotion perception has focused on facial expressions, and findings from this modality have come to dominate our thinking about other modalities. Here, we examine emotion perception through a wider lens by comparing facial with vocal and tactile processing. We review stimulus characteristics and ensuing behavioral and brain responses, and show that audition and touch do not simply duplicate visual mechanisms. Each modality provides a distinct input channel and engages partly non-overlapping neuroanatomical systems with different processing specializations (e.g., specific emotions versus affect). Moreover, processing of signals across the different modalities converges, first into multi- and later into amodal representations that enable holistic emotion judgments.

Keywords

Nonverbal Emotions – Moving from a Unimodal to a Multimodal Perspective

Emotion perception plays a ubiquitous role in human interactions and is hence of interest to a range of disciplines, including psychology, psychiatry, and social neuroscience. Yet, its study has been dominated by facial emotions, with other modalities explored less frequently and often anchored within a framework derived from what we know about vision. Here, we take a step back and put facial emotions on a par with vocal and tactile emotions. Like a frightened face, a shaking voice or a cold grip can be meaningfully interpreted as emotional by us. We will first explore these three modalities in terms of signal properties and brain processes underpinning unimodal **perception** (see Glossary). We will then examine how signals from different channels converge into a holistic understanding of another's **feelings**.

Correspondence: radolphs@caltech.edu, schirmer@nus.edu.sg.

Throughout, we will address the question whether emotion perception is the same or different across modalities.

Sensory Modalities for Emotion Expression

There is vigorous debate about what exactly individuals can express nonverbally. There are hotly debated open questions about whether people are in an objective sense "expressing emotions" as opposed to engaging in more strategic social communication, whether they express discrete emotions and which ones, and whether their expressions are culturally universal. For practical reasons, we ignore these debates here and simply assume that people do express and perceive what we usually call **emotions** (Box 1).

Box 1

How Are Emotion Expression and Perception Defined?

Emotional expressions are motivated behaviors co-opted to be perceived as communicative signals, but that may or may not be intended as such. Darwin was the first to propose three causal origins for expressions [85]. One of them involves an immediate benefit to the expressive individual; for example, increasing one's apparent body size serves to intimidate an opponent. Other expressions, Darwin suggested, effectively communicate the polar opposite; for example, lowering one's body-frame to signal submission instead of aggression. Lastly, Darwin held that some expressions, such as the trembling in fear, are mere vestigial by-products that may not serve a useful role. Although many details of Darwin's proposal are debatable, there is evidence that some facial expressions exhibit useful functions (e.g., widening of the eyes to maximize the visual field during fear) [86] and/or to effectively manipulate the behavior of perceivers [87].

Perception, recognition, and categorization are probed with distinct tasks. Emotion perception includes early detection and discrimination, and is often conceptualized as providing input to the richer inferences that we make about people's internal states and the retrieval of associated conceptual knowledge ("recognition"), the sorting into emotion categories often defined by words in a particular language ("categorization"), as well as our own emotional responses that can be evoked (e.g., empathy) [88]. Building on earlier work suggesting that perceptual representations feed into systems for judging emotions [89], neuroimaging findings point to specific nodes in regions such as the posterior STS that may mediate between early emotion perception and mental state inference [90]. Similarly, regions of right somatosensory cortex within which lesions [20] or brain stimulation [91] can alter multimodal emotion recognition [92,93], show multivoxel activations in fMRI that correlate with self-reported emotional experiences of the perceiver [30]. There are thus rich connections between early perceptual processing, later mental state inference, and emotion induction in the perceiver. Most studies do not distinguish between these stages, an issue that also confounds the interpretation of putative top-down effects of cognition on perception [94]. A fuller picture of the flow of information from perceptual to inferential processes will also benefit from techniques like

magnetoencephalography (MEG), which combines good spatial resolution with excellent temporal resolution [95,96].

Facial expressions have been studied in the most detail, possibly because they seem most apparent in everyday life and their controlled presentation in an experimental setting is relatively easy. Facial expressions in humans depend on 17 facial muscle pairs that we share fully with great apes and partially with some other species (Box 2). The neural control of these muscles depends on a network of cortical and subcortical structures with neuroanatomical and functional specialization [1]. For instance, the medial subdivision of the facial nucleus, which moves our ears, is relatively underdeveloped in humans compared to mammals that can move their ears, whereas the lateral subdivision, which moves our mouth, is exceptionally well developed [2].

Box 2

Emotion Expression and Perception Exist in Non-Human Animals

Following Darwin's [85] description of homologues in the emotional behaviors of mammals, there are now quantitative methods similar to the human FACS for describing the facial expressions of other species. The application of these methods has revealed expression similarities. For example, humans share the zygomaticus major with other primates and dogs who, like us, use it to retract mouth-corners and expose teeth [97,98]. Moreover, across these species zygomaticus activity together with the dropping of the jaw contribute to a joyful display termed the "play face" (Figure I).

Importantly, species similarities extend to other expressive channels including the voice and touch. Human fearful screams, as an example, share many acoustic properties (e.g., deterministic chaos, high intensity, high pitch) with the screams of other primates and mammals in general [99]. Additionally, in humans as in many other social species, affiliative emotions trigger close physical contact and gentle touch [14].

Species similarities in expression are matched by similarities in perception as evidenced by both behavioral and brain responses. In monkeys encountering a novel object, the concurrent display of positive and negative facial expressions bias approach and avoidance behaviors, respectively [100]. Tree shrews respond differentially to the same call produced with high and low affective intensity [101]. Zebrafish show a faster reduction in fear behaviors and cortisol following stress when a water current stroked their lateral lines during recovery [102].

Corresponding to these perceptual effects are neural systems that bear resemblance to those identified in humans. In nonhuman primates, inferior and superior temporal regions are likely homologues to the human face and voice areas, respectively, and are linked to a network for crossmodal convergence including prefrontal cortex and STS [103,104]. Dogs, like primates, show face sensitivity in their inferior temporal lobes [105]. Moreover, their brains respond differently to neutral (e.g., "such") and praise words (e.g., "clever") depending on the speaker's voice: only a positive word in combination with a positive voice activates reward-related brain regions [106]. Lastly, emotion perception seems relatively lateralized to the right hemisphere in a wide range of species [107]. In

sum, available evidence speaks for homology in the neural systems subserving emotion expression and perception, at least across mammals. Nonetheless, figuring out exactly how these signals influence social behaviour will require substantially more ethological research.
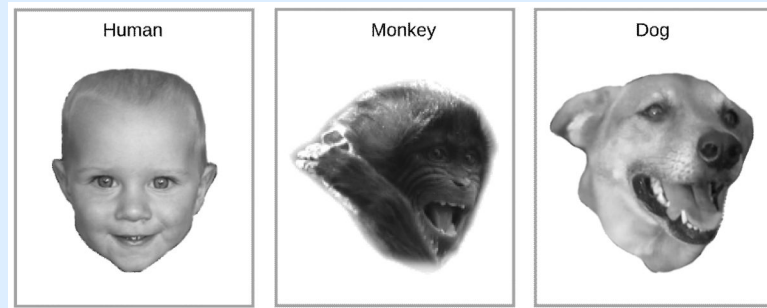

Figure I.
Positive facial affect in humans, apes and dogs. Adapted with permission from Schirmer and colleagues (2013) as well as Palagi and Mancini [110].

Quantification of facial muscle movements and their composition into emotional expressions is provided by the Facial Action Coding System (FACS) [3]. Different expressions can be characterized by linear combinations of action units (basic sets of muscle movements), whereby a single unit can participate in multiple emotional expressions (e.g., brow lowerer, Figure 1). Displays that are typically categorized above chance include happiness, sadness, fear, anger and disgust [4]. However, defining accuracy for this as well as other expressive modalities depends on methodological choices, including the posers' acting ability or expressive disposition, the set of emotion categories to which the stimulus is typically matched in the task, as well as cultural points of reference [5,6]. Although we do not cover it here, emotion perception from body postures is another relevant visual channel, perhaps more important in certain animals like cats and dogs rather than humans, and one which interacts with other modes of expression [7].

In addition to faces, voices are an important modality for emotion communication. They emerge from the vocal system via complex movements that involve the abdomen, larynx, tongue, as well as cavities along the vocal tract that filter airwaves and produce resonance effects. Voices convey inborn and learned emotion signals (e.g., screams, sobs, laughs) and their sound may be affected by emotion-induced physiological changes of the vocal system including changes in breathing and muscle tone. Vocal expressions in the context of speech are referred to as speech melody (i.e., prosody) and can be dissociated from verbal content [8].

Sound characteristics can be quantified with a range of acoustic parameters including loudness, fundamental frequency (i.e., melody), and **voice quality** (e.g., roughness, Figure 1). Some of these parameters show a robust relationship with perceived speaker **affect**. For instance, loudness and fundamental frequency each correlate positively with arousal [9,10]. The relationship between vocal acoustics and specific emotions is a bit more tenuous, but

some evidence suggests that listeners rely on voice quality, individually and in combination with other acoustic parameters [11]. Although emotion recognition can be fairly accurate when listeners chose from a limited set of emotion categories, agreement drops significantly as more categories become available. Moreover, fewer emotions can be perceived from the voice than from the face [12].

By contrast to the perception of faces and voices, the perception of touch requires direct physical contact. In the context of benign human interaction, touch is typically an affectively positive stimulus perceived as pleasurable and promoting bonding and trust [13,14]. Contributing to these effects is a special tactile receptor, the **C-tactile (CT) afferent**, found in non-glabrous skin. CT-afferents fire maximally to skin-temperature touch [15] and to light-pressure stroking of 3 to 10 cm per second [16]. Because their firing rate correlates positively with the subjective pleasure that people experience from touch, they are thought to play a central role in the social and affective aspects of somatosensation [16].

In addition to the intrinsically rewarding aspects of touch, its specific forms can communicate individual emotions. Based on the body part that is being stimulated, the tactile action (e.g., pushing, stroking), its attributes (e.g., force, duration), and the context (e.g., touch from a lover or a stranger) people can infer a range of emotions. For instance, American college students can correctly categorize anger, fear, disgust, love, gratitude, sympathy, happiness, and sadness from such stimuli when presented in a controlled setting [17]. While relatively unexplored, such emotion categorization very likely involves higher-order inferences and culturally acquired knowledge about social rules and metaphors.

Neural Systems for Perceiving Emotions

The neural systems underpinning the perception of emotions have been studied with a range of tasks and techniques. The following review emphasizes approaches that were used frequently and consistently across modalities and, hence, support a systematic comparison between face, voice and touch. With respect to tasks, those contrasting nonverbal expressions with control stimuli (e.g., face-house) as well emotional with neutral expressions were selected as most relevant in our review. With respect to techniques, lesion studies, simple contrasts or pattern classifications of **fMRI** data, as well as event-related potentials (**ERPs**) of the **EEG** were selected.

Facial Emotions

The cortical system for primate face processing encompasses a mosaic of patches in temporal, occipital, parietal, and frontal cortices that collectively process particular features of faces and generate a distributed perceptual representation [18]. Of particular prominence in human fMRI studies is an area in the inferior temporal cortex that is more strongly activated by faces than non-face objects, the fusiform face area (FFA, Figure 2) [19]. The cortical face network is complemented by a key subcortical structure, the **amygdala**.

Insight into neural components that might be more specific for judging emotions from faces, rather than subserving face processing in general, has come from several approaches. Earlier lesion studies pointed to regions in right ventro-parietal and posterior-temporal cortices,

including the angular and supramarginal gyri, and even components of somatosensory cortex [20]. However, these regions may be more specifically involved in fine-grained inferences about the emotion from faces that depend on a particular **embodiment** mechanism, such as representing how the emotion would feel through real or imagined mimicry [21]. Neuroimaging revealed that facial emotions modulate amygdala activity as well as activity in other components of the face processing system [22]. Single amygdala neurons respond to facial emotions [23], in particular expressions of the eye region [24]. Whereas cortical face processing [25] and its emotion modulation [26] appear preferentially lateralized to the right hemisphere, the lateralization of subcortical emotion effects is still unclear.

There is some evidence that expressions of different emotions recruit somewhat segregated face processing circuits. This was demonstrated with **multi-voxel pattern analysis** (MVPA) within FFA and primary visual cortex [27], medial prefrontal cortex (mPFC) [28], and superior temporal sulcus (STS) [29]. Additionally, an intriguing recent study found that multivoxel decoding of individual emotions could be achieved from somatosensory cortex. The accuracy of such decoding showed a somatotopic pattern: for example, the perception of fear, which is distinguished by wide eyes, could be decoded best from the region of somatosensory cortex that represents the eyes [30]. Case studies of rare lesion patients have provided evidence for emotion differentiation in subcortical structures as well. For example, recognition of fear and disgust can be impaired by lesions to the amygdala [31,32] and the basal ganglia [33], respectively. However, neither of these findings is well corroborated by fMRI studies and neither clearly separates perception from other post-perceptual biases.

Complementing the excellent anatomical resolution of fMRI are techniques, such as the **EEG**, that yield excellent temporal resolution. **ERPs**, which are derived from the EEG by time-locked averaging in a particular task, are characterized by a series of positive and negative deflections of which some are modulated by emotion. Of particular relevance in the context of faces is the N170, a negativity peaking about 170 ms following stimulus onset over temporal electrodes that is specific to faces and for which sources lie in the FFA [34]. This negativity [35], but also a later positive potential that is nonspecific for faces [36], are greater to emotional as compared to neutral expressions, especially if they are highly arousing (Figure 3). Furthermore, some electrophysiological responses to facial emotions occur intracranially with latencies barely exceeding 100 ms even in traditionally "high-level" regions such as prefrontal cortex [37], arguing for multiple streams of visual processing that contribute to emotion perception.

In particular, there is a long-standing debate about the possibility that the amygdala receives coarse (low spatial frequency) information about emotional faces through a rapid route via the superior colliculus [38]. One recent intracranial recording study found evidence supporting this idea from field potentials recorded from the amygdala of neurosurgical patients [39], even though action potentials of amygdala neurons to faces have very long latencies inconsistent with this view. Some information about facial emotions may thus be represented already very early in the visual system, and may bias subsequent processing [40].

Vocal Emotions

Similar to the better known visual system, the auditory system comprises two pathways originating from different sections within the respective sensory nucleus of the thalamus, the medial geniculate thalamus. One pathway, referred to as lemniscal, has a strong preference for auditory signals, is **tonotopically** organized, relays fast signal aspects and targets primary auditory cortex (BA41, core). The second, non-lemniscal pathway is multimodal, weakly tonotopic, prefers slow signal aspects, and targets predominantly secondary auditory cortex (BA42, belt and parabelt) [41].

Voices recruit both lemniscal and non-lemniscal pathways [41]. When compared with non-vocal sounds, they activate bilateral secondary auditory cortex together with surrounding superior temporal cortex more strongly [42,43, Figure 2]. These mid-temporal regions are often referred to as "voice areas" by analogy to the FFA. Because they are also recruited by other complex sounds (e.g., music, environmental sounds), they may not be voice-specific. As for faces [44], some of the differential responses to vocalizations may simply result from them being more relevant, practiced, or attention capturing [45].

Emotional vocalizations produce greater activation in voice areas than do neutral vocalizations and this difference is relatively lateralized to the right hemisphere [46,47]. Additionally, there are reports of amygdala lesions impairing vocal emotion recognition and altering associated cortical processes [48,49]. However, PET or fMRI studies in healthy populations provide only weak support for the amygdala's role: greater amygdala activation to emotional than neutral voices is found typically only with more liberal statistical thresholding. Possibly, multimodal neurons in the amygdala establish associations with face processing [50,51] but contribute to unimodal voice perception in a fairly subtle way.

As for facial emotions, there is some indication that brain representations diverge for different vocal emotions. Using MVPA, multiple vocal emotions could be categorized based on their specific activation patterns in the temporal voice areas [52] and within mPFC and posterior superior temporal cortex [28]. Additionally, brain lesions have been reported to produce emotion-specific deficits. Similar to facial disgust, vocal disgust may be perceived less accurately following insular damage [53], whereas vocal fear and anger may be perceived less accurately following amygdala damage [49], although these lesion studies will require larger samples for corroboration.

The timecourse of voice perception has been investigated with ERPs. Compared to non-vocal sounds, vocalizations enhance a positivity around 200 ms following stimulus onset with sources in the temporal voice areas [54]. Moreover, within that same time range, ERPs differentiate between emotional and neutral voices [55,56, Figure 3] with amplitude modulations predicting changes in the affective connotation of concurrent verbal content [57]. Specifically, neutral words heard in an emotional voice are subsequently rated as more emotional the larger their positive amplitude in the ERP. Together, this evidence implies that affective information from the voice is available within 200 ms following voice onset. Furthermore, as for faces, this temporally coincides with the processing of other voice information, potentially providing an early bias for emotions.

Emotional Touch

Touch constitutes one amongst several somatosensory channels that include the sensation of temperature, pain, and the position of one's body in space [58]. CT-afferents as well as a variety of mechanoreceptors collectively referred to as Aβ-fibers convey mechanical stimulation from the skin to the brain. Whereas CT signals travel slowly, via unmyelinated fibers primarily along the spinothalamic pathway, Aβ signals travel quickly via myelinated fibers primarily along the posterior-column medial lemniscus pathway. Moreover, whereas CT projections reach the posterior insula directly from the somatosensory thalamus, Aβ projections travel from the thalamus first to primary and secondary somatosensory cortex. Compared to Aβ processing, CT processing has more in common with **interoception** – it is more affective and less spatially discriminative [58]. Rare peripheral demyelinating diseases can abolish Aβ touch while sparing CT touch. Such patients are impaired at localizing touch on their body, but can still perceive it as pleasant, and selectively activate the insula [59].

To date, neuroimaging studies on tactile emotions have focused on perceived pleasure rather than communicative function. A recent meta-analysis identified the right posterior insula as a structure that is more frequently activated by pleasurable as compared to neutral touch [60, Figure 2]. Studies focusing specifically on the CT system additionally highlight the right posterior STS [61–63] and somewhat less consistently the right orbito-frontal [64–66] and pre-motor/motor areas [62,65,67]. There is some evidence that primary somatosensory cortex contributes to the pleasure from touch: a caress activates this region differentially depending on whether participants believe it to come from a person of the same or the opposite sex [68]. This finding may be an example of higher-order attributional feedback, since a combined fMRI/TMS study using a mechanical touch device found that pleasure ratings of CT-appropriate stroking correlated with activity outside the primary somatosensory cortex and that **TMS** stimulation to this area altered the perception of touch intensity but not pleasure [67].

Again, insights into the timecourse of emotional touch perception have been sought with electrophysiological measures. Several decades of research have established a somatosensory ERP to tactile stimulation that resembles the ERP for stimuli from other modalities and shows prominent deflections within the first 100 ms following touch onset [69,70]. So far, however, efforts have focused on the effects of benign or painful stimulation to the finger tips, which are dense with ordinary Aβ-fibers but lack CT-afferents. Only one study explored CT appropriate touch and failed to identify a clear somatosensory potential [71]. The ERP associated with light pressure to the forearm caused by an inflating pressure cuff or the hand of a friend showed no visible deflections. Perhaps the gradual increase in pressure could not be picked up with the ERP because it was insufficiently time-locked. Additionally, cortical processing – the main contributor to the ERP – may be limited for CT touch, especially if it is task-irrelevant.

Modality Similarities, Differences and Convergence

Modality Similarities and Differences

To date, research on emotion perception has emphasized the face. Moreover, insights from the face have served as a general guide to search for and establish analogies with other modalities [72,73]. And indeed such analogies exist. Vision, audition, and touch each have slow and fast processing pathways that project to specialized regions of sensory cortex, with upstream regions generally responding more robustly to social over non-social signals, especially if they are emotional (Table 1). Additionally, across these three modalities, contrasts between social and non-social stimuli on the one hand, and between emotional and neutral expressions on the other hand, activate overlapping brain regions with a right hemisphere bias and a similar timecourse.

Although we have emphasized central mechanisms of perception, there are also important perceptual differences arising from the nature of the stimulus itself, and from the way that stimulus is produced. One example is the temporal dimension of nonverbal expression [74]. Although all expressions extend in time, faces and body postures can be discerned from a single snapshot, a fact that is being leveraged in most visual studies by presenting still images. In contrast, the investigation of vocalizations and touch necessitates a dynamic stimulus for which receivers must integrate information over time.

Second, the different modalities have different physical constraints imposing different demand characteristics on emotion perception. The face with its many muscles can be minutely controlled to produce a wide range of expressions that perceivers must differentiate. Moreover, the results of this process map onto fairly specific emotion categories rather than simple affect [75]. By comparison, typical interpersonal touch is more limited in variance with regards to temperature, speed, and pressure. As such associated processing demands are comparatively low and perceptual representation more likely to carry simple affective qualities as opposed to information about more differentiated emotions. Exceptions occur when the significance of touch is shaped by additional factors, such as where it is placed on the body or its cultural significance. This may result in context-dependent associations with specific emotions.

Third, emotional expressions show interdependencies among some modalities but not others. For example, facial movements inadvertently shape the mouth region and thereby impact vocal acoustics. In turn, vocalizing by engaging the mouth alters concurrent facial displays. As a consequence, facial and vocal expressions are not independent and influence perceptual processes in a coordinated manner. Yet both have little or no impact on tactile modes of communication and vice versa.

Last, there is evidence for modality-specificity in the neural pathways underpinning perception. Signal representations in primary auditory cortex are significantly more processed than those in primary visual cortex [73]. Pleasurable human touch is known to bypass primary sensory cortex to produce emotions independently of a conscious tactile sensation [59]. In contrast, similar subcortical routes for sound and sight are still being debated. Additionally, to the extent that body representations are involved when we make

emotion judgments, touch may evoke emotions fairly immediately due to its interoceptive quality, whereas the voice and face may do so more indirectly via further inferences and **embodiment**.

Given all these differences, it should come as no surprise that the visual, auditory, and tactile senses each offer particular strengths and limitations when it comes to emotion perception. As a consequence, emerging unimodal representations are only partially redundant, necessitating modality convergence for a holistic and ecological understanding of emotion.

Convergence of the Senses

The mechanisms of modality convergence have long been investigated through behavioral, fMRI, and ERP studies [111]. A clear behavioral finding is that compared to unimodal presentations (e.g., face only), multimodal presentations (e.g., face and voice) yield faster and more accurate emotion judgments [75]. Similarly, fMRI contrasts show greater activation to multi- as compared to unimodal expressions in a range of brain regions including thalamus, STS, FFA, insula and amygdala as well as higher order association areas like lateral and medial PFC and the posterior cingulate [75, Figure 3]. Increased activation in the pSTS, a known convergence zone for expressions from the different senses, is associated with increased functional connectivity between this region and unimodal visual and auditory cortex [76].

Corroborating these results, MVPA successfully categorized emotions across different modalities in a subset of the areas that were found to be more activated for multi- as compared to unimodal stimuli. For example, emotional faces, voices, and body postures could all be classified from activation in STS and medial PFC [28]. Emotional faces and fractals for which participants had learned to associate an emotion could be classified from activation in medial PFC and posterior cingulate [77].

Lastly, ERP research revealed that, compared to unimodal stimulation, multimodal stimulation produces larger amplitude differences between emotional and neutral conditions. Moreover, such multi-modal effects have been reported within 100 ms following stimulus onset [78], whereas comparable unimodal effects emerge typically at 170 ms (faces) or 200 ms (voices) only.

These data indicate that multimodal integration is not simply a late process that occurs after the individual modalities have been analyzed [111]. Instead, it is supported by multimodal neurons at some of the earliest processing stages, including the superior colliculus (SC) [79,80] and the thalamus [81], and increases in strength along cortical pathways [40]. In more detail, the SC, a midbrain structure, divides into superficial layers that are exclusively visual, and deeper layers that combine visual, auditory, and somatosensory input. Reciprocal connections between these layers and their onward projections to the thalamus likely support early aspects of multimodal processing such as temporal binding and cross-modal enhancement (e.g., reducing a unimodal detection threshold) [79,80].

Multimodal processing at the telencephalic level may be two-fold. Regions like the FFA, amygdala, middle STS, and posterior insula may support largely perceptual representations.

Here, different physical cues may be "assembled" and mapped onto a stored template previously referred to as an "emotional gestalt" [82]. At a more advanced stage of processing, regions like the PFC, the posterior cingulate, and posterior STS may support amodal, conceptual representations that involve a modality-nonspecific abstract code [82,83]. Although further research is needed to better specify the computations at subcortical and cortical levels, available evidence suggests that multimodal representations emerge incrementally (Figure 4).

Resultant benefits of convergence for the individual are manifold. As mentioned above, multimodal convergence ensures that distributed emotion signals are being integrated into a holistic percept, which is also generally more reliable. Additionally, convergence underpins temporal binding across unimodal processing streams and the formation of temporal predictions concerning the behavior of interaction partners [74]. Affectively congruous expressions are processed more readily than incongruous ones [84], an effect perhaps subserved by Hebbian associations. Furthermore, multimodal incongruity may provide clues to deception, as when a smiling face belies a more aggressive tone of voice. Learned multimodal associations may also help interpret the personal expressive styles of particular individuals.

Concluding Remarks

Social interactions elicit spontaneous and strategic expressions that we commonly classify as emotions, and that profoundly influence a perceiver's mental state and ongoing behavior [13,57,71]. Neuroscience investigations have been uneven, with a historical overemphasis on facial signals. Although many questions remain (see Outstanding Questions), enough evidence has accumulated to characterize mechanisms both unique and shared across modalities. Furthermore, although each channel engages modality-specific sensory systems, the perceptual representations emerging from these systems are integrated very early. Uni- and multimodal streams feed back to earlier levels as well as project onward to perceptual and conceptual nodes distributed throughout the cortex. To ultimately understand how emotion signals guide our social interactions, future studies should adopt a rich approach that cuts across methods and temporal scales, that acknowledges the ecological validity of dynamic stimuli, and that involves not one but multiple sensory modalities.

Outstanding Questions

Do individual emotion categories activate distinct neural regions?

While there is good evidence that MVPA within regions can distinguish emotion categories, the role of entire neural structures is less clear. For instance, while lesion studies suggest the amygdala may be necessary specifically for fear recognition, this is not corroborated by fMRI studies.

What are the relations between emotion expression, perception, and experience?

It is still open whether emotion recognition necessitates an emotional experience and whether such experience is causally involved in the recognition. Similarly, emotion

---

perception often causes mimicry in the perceiver, which is in turn perceived; such coupling likely plays an important role in how emotion communication unfolds in time.

What are good tasks for studying spontaneous and intentional emotion recognition?

The tasks used to study emotion recognition are typically highly artificial (e.g., change detection, forced-choice categorizing). It will be important to develop paradigms that more closely approximate how we engage with emotion expressions in real life.

Do current insights hold when stimuli are made more ecologically valid?

In an effort to avoid confounds, many studies use impoverished stimuli that are devoid of context. For example, many visual studies use still images and many tactile studies use a tactile device rather than actual human touch.

How prominent are individual differences?

Most studies to date have explored average responses across all study participants. However, there is accumulating evidence that individual factors like sex, age, personality and culture influence both emotion expression and perception.

What social signals qualify as expressions of emotions?

We make a lot of nonverbal sounds and move our faces a lot. Are all of these "emotional?" We would suggest reserving that term for only a subset, but considerably more theoretical work is needed to clarify where to draw the boundary.

Glossary

Affect

often assumed to be a precursor or prerequisite for a more differentiated emotion, affect is typically described as two-dimensional, varying in valence (pleasant to unpleasant) and arousal (relaxed to excited). While applied primarily to the structure of feelings, it can also be applied to behavioural consequences of emotion states (e.g., approach-withdrawal).

Amygdala

an almond-shaped nucleus situated in the anterior aspect of the medial temporal lobe. It comprises several sub-nuclei and contributes to a diverse set of functions including sexuality, memory, and emotions. In the context of emotion perception, the amygdala has been implicated particularly in the recognition of fear from facial expressions.

CT-afferent

a tactile receptor found in non-glabrous (i.e., hairy) skin that conveys pressure, speed and temperature information via slowly conducting C fibers – a class of fibers that typically enables pain. However, CT responses are tuned to the characteristics of benign human touch.

FMRI

functional magnetic resonance imaging detects small changes in blood oxygenation in the brain to infer bulk changes in neuronal activity. It enables the investigation of stimulus- or

task-dependent regional changes in brain activation, as well as the investigation of functional connectivity between regions.

EEG

electroencephalogram, a technique that records the bulk electrical activity of cortical neurons with electrodes placed on the scalp. It provides much better temporal resolution than fMRI (milliseconds compared to seconds) but poorer spatial resolution.

Embodiment

refers to the engagement of one's own body, or representations thereof, in cognition. For instance, embodiment of observed actions is hypothesized to facilitate their comprehension: facial expressions are recognized more accurately when individuals can spontaneously mimic the expression they see.

Emotions

in this review, emotions are taken to be complex neurobiological states elicited by situations relevant for an individual's current or prospective needs. They motivate and coordinate cognitions and behaviors that help fulfil these needs.

ERP

event-related potentials are derived by time-locked averaging of the EEG, typically in response to a specific stimulus. They comprise a series of positive and negative deflections referred to as ERP components.

Feelings

the conscious experiences of emotions, typically measured with verbal report in humans.

Interoception

the perception of one's internal bodily state (e.g., heart rate).

Multi-voxel pattern analysis (MVPA)

a multivariate machine learning approach used for fMRI data analysis. It analyzes the response pattern of multiple individual voxels within a specified brain region. In a first step, a classifier learns to distinguish the repertoire of patterns from two conditions. In a second step, the trained classifier is applied to novel data (a novel pattern) and its accuracy in distinguishing between the two conditions is established. The approach capitalizes on the assumption that information in the brain is represented in a high-dimensional format – in variance over patterns, rather than mere amplitude of response.

Perception

initial receptive processing encompassing abilities such as detection and discrimination, often thought of as providing the input representations for subsequent recognition and categorization.

Tonotopy

the spatial organization of neurons into maps (topography) so that they are organized according to their tuning to sound frequency.

Transcranial magnetic stimulation (TMS)

experimental manipulation of neuronal activity over a region of cortex achieved with magnetic field pulses that induce local currents in the brain.

Voice quality

a composite of vocal characteristics that are shaped by vocal tract configuration and laryngeal anatomy. They arise in part from resonance properties of the vocal system and create sound impressions such as hoarse, nasal or breathy. (http://www.ncvs.org/ncvs/tutorials/voiceprod/tutorial/quality.html).

References

1. Müri RM. Cortical control of facial expression. J Comp Neurol. 2016; 524:1578–1585. [PubMed: 26418049]

2. Sherwood CC, et al. Evolution of the brainstem orofacial motor system in primates: a comparative study of trigeminal, facial, and hypoglossal nuclei. J Hum Evol. 2005; 48:45–84. [PubMed: 15656936]

3. Ekman P, Friesen WV. Measuring facial movement. Environ Psychol Nonverbal Behav. 1976; 1:56–75.

4. Jack RE, et al. Facial expressions of emotion are not culturally universal. Proc Natl Acad Sci. 2012; 109:7241–7244. [PubMed: 22509011]

5. Gendron M, et al. Perceptions of emotion from facial expressions are not culturally universal: evidence from a remote culture. Emot Wash DC. 2014; 14:251–262.

6. Gendron M, et al. Cultural Relativity in Perceiving Emotion From Vocalizations. Psychol Sci. 2014; 25:911–920. [PubMed: 24501109]

7. Enea V, Iancu S. Processing emotional body expressions: state-of-the-art. Soc Neurosci. 2016; 11:495–506. [PubMed: 26513592]

8. Banse R, Scherer KR. Acoustic profiles in vocal emotion expression. J Pers Soc Psychol. 1996; 70:614–636. [PubMed: 8851745]

9. Bänziger T, et al. Path Models of Vocal Emotion Communication. PLOS ONE. 2015; 10:e0136675. [PubMed: 26325076]

10. Laukka P, et al. A dimensional approach to vocal expression of emotion. Cogn Emot. 2005; 19:633–653.

11. Birkholz P, et al. The contribution of phonation type to the perception of vocal emotions in German: An articulatory synthesis study. J Acoust Soc Am. 2015; 137:1503–1512. [PubMed: 25786961]

12. Cordaro DT, et al. The voice conveys emotion in ten globalized cultures and one remote village in Bhutan. Emotion. 2016; 16:117–128. [PubMed: 26389648]

13. Brauer J, et al. Frequency of maternal touch predicts resting activity and connectivity of the developing social brain. Cereb Cortex. 2016; doi: 10.1093/cercor/bhw137

14. Dunbar RIM. The social role of touch in humans and primates: Behavioural function and neurobiological mechanisms. Neurosci Biobehav Rev. 2010; 34:260–268. [PubMed: 18662717]

15. Ackerley R, et al. Human C-Tactile Afferents Are Tuned to the Temperature of a Skin-Stroking Caress. J Neurosci. 2014; 34:2879–2883. [PubMed: 24553929]

16. Löken LS, et al. Coding of pleasant touch by unmyelinated afferents in humans. Nat Neurosci. 2009; 12:547–548. [PubMed: 19363489]

17. Hertenstein MJ, et al. The communication of emotion via touch. Emot Wash DC. 2009; 9:566–573.

18. Tsao DY, Livingstone MS. Mechanisms of face perception. Annu Rev Neurosci. 2008; 31:411–437. [PubMed: 18558862]

19. Kanwisher N, et al. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci Off J Soc Neurosci. 1997; 17:4302–4311.

20. Adolphs R, et al. A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. J Neurosci Off J Soc Neurosci. 2000; 20:2683–2690.

21. Keysers C, Gazzola V. Integrating simulation and theory of mind: from self to social cognition. Trends Cogn Sci. 2007; 11:194–196. [PubMed: 17344090]

22. Vuilleumier P, Pourtois G. Distributed and interactive brain mechanisms during emotion face perception: evidence from functional neuroimaging. Neuropsychologia. 2007; 45:174–194. [PubMed: 16854439]

23. Wang S, et al. Neurons in the human amygdala selective for perceived emotion. Proc Natl Acad Sci U S A. 2014; 111:E3110–3119. [PubMed: 24982200]

24. Rutishauser U, et al. The primate amygdala in social perception – insights from electrophysiological recordings and stimulation. Trends Neurosci. 2015; 38:295–306. [PubMed: 25847686]

25. De Winter FL, et al. Lateralization for dynamic facial expressions in human superior temporal sulcus. NeuroImage. 2015; 106:340–352. [PubMed: 25463458]

26. Innes BR, et al. A leftward bias however you look at it: Revisiting the emotional chimeric face task as a tool for measuring emotion lateralization. Laterality. 2015; doi: 10.1080/1357650X. 2015.1117095

27. Skerry AE, Saxe R. Neural Representations of Emotion Are Organized around Abstract Event Features. Curr Biol. 2015; 25:1945–1954. [PubMed: 26212878]

28. Peelen MV, et al. Supramodal Representations of Perceived Emotions in the Human Brain. J Neurosci. 2010; 30:10127–10134. [PubMed: 20668196]

29. Wegrzyn M, et al. Investigating the brain basis of facial expression perception using multi-voxel pattern analysis. Cortex. 2015; 69:131–140. [PubMed: 26046623]

30. Kragel PA, LaBar KS. Multivariate neural biomarkers of emotional states are categorically distinct. Soc Cogn Affect Neurosci. 2015; 10:1437–1448. [PubMed: 25813790]

31. Adolphs R, Tranel D. Intact recognition of emotional prosody following amygdala damage. Neuropsychologia. 1999; 37:1285–1292. [PubMed: 10530728]

32. Adolphs R. What does the amygdala contribute to social cognition? Ann N Y Acad Sci. 2010; 1191:42–61. [PubMed: 20392275]

33. Calder AJ, et al. Neuropsychology of fear and loathing. Nat Rev Neurosci. 2001; 2:352–363. [PubMed: 11331919]

34. Bentin S, et al. Electrophysiological Studies of Face Perception in Humans. J Cogn Neurosci. 1996; 8:551–565. [PubMed: 20740065]

35. Hinojosa JA, et al. N170 sensitivity to facial expression: A meta-analysis. Neurosci Biobehav Rev. 2015; 55:498–509. [PubMed: 26067902]

36. Suess F, et al. Perceiving emotions in neutral faces: expression processing is biased by affective person knowledge. Soc Cogn Affect Neurosci. 2015; 10:531–536. [PubMed: 24948155]

37. Kawasaki H, et al. Single-neuron responses to emotional visual stimuli recorded in human ventral prefrontal cortex. Nat Neurosci. 2001; 4:15–16. [PubMed: 11135639]

38. Pessoa L, Adolphs R. Emotion processing and the amygdala: from a "low road" to "many roads" of evaluating biological significance. Nat Rev Neurosci. 2010; 11:773–783. [PubMed: 20959860]

39. Méndez-Bértolo C, et al. A fast pathway for fear in human amygdala. Nat Neurosci. 2016; 19:1041–1049. [PubMed: 27294508]

40. Kveraga K, et al. Magnocellular projections as the trigger of top-down facilitation in recognition. J Neurosci Off J Soc Neurosci. 2007; 27:13232–13240.

41. Kraus N, White-Schwoch T. Unraveling the Biology of Auditory Learning: A Cognitive–Sensorimotor–Reward Framework. Trends Cogn Sci. 2015; 19:642–654. [PubMed: 26454481]

42. Pernet CR, et al. The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. NeuroImage. 2015; 119:164–174. [PubMed: 26116964]

43. Talkington WJ, et al. Humans Mimicking Animals: A Cortical Hierarchy for Human Vocal Communication Sounds. J Neurosci. 2012; 32:8084–8093. [PubMed: 22674283]

44. McGugin RW, et al. High-resolution imaging of expertise reveals reliable object selectivity in the fusiform face area related to perceptual performance. Proc Natl Acad Sci U S A. 2012; 109:17063–17068. [PubMed: 23027970]

45. Schirmer A, et al. On the spatial organization of sound processing in the human temporal lobe: A meta-analysis. NeuroImage. 2012; 63:137–147. [PubMed: 22732561]

46. Brück C, et al. Impact of personality on the cerebral processing of emotional prosody. NeuroImage. 2011; 58:259–268. [PubMed: 21689767]

47. Mothes-Lasch M, et al. Visual Attention Modulates Brain Activation to Angry Voices. J Neurosci. 2011; 31:9594–9598. [PubMed: 21715624]

48. Frühholz S, et al. Asymmetrical effects of unilateral right or left amygdala damage on auditory cortical processing of vocal emotions. Proc Natl Acad Sci U S A. 2015; 112:1583–1588. [PubMed: 25605886]

49. Scott SK, et al. Impaired auditory recognition of fear and anger following bilateral amygdala lesions. Nature. 1997; 385:254–257. [PubMed: 9000073]

50. Aubé W, et al. Fear across the senses: brain responses to music, vocalizations and facial expressions. Soc Cogn Affect Neurosci. 2015; 10:399–407. [PubMed: 24795437]

51. Klasen M, et al. Supramodal Representation of Emotions. J Neurosci. 2011; 31:13635–13643. [PubMed: 21940454]

52. Ethofer T, et al. Decoding of Emotional Information in Voice-Sensitive Cortices. Curr Biol. 2009; 19:1028–1033. [PubMed: 19446457]

53. Calder AJ, et al. Impaired recognition and experience of disgust following brain injury. Nat Neurosci. 2000; 3:1077–1078. [PubMed: 11036262]

54. Capilla A, et al. The early spatio-temporal correlates and task independence of cerebral voice processing studied with MEG. Cereb Cortex N Y N 1991. 2013; 23:1388–1395.

55. Jiang X, Pell MD. On how the brain decodes vocal cues about speaker confidence. Cortex. 2015; 66:9–34. [PubMed: 25796072]

56. Paulmann S, Kotz SA. Early emotional prosody perception based on different speaker voices. Neuroreport. 2008; 19:209–213. [PubMed: 18185110]

57. Schirmer A, et al. Vocal emotions influence verbal memory: Neural correlates and interindividual differences. Cogn Affect Behav Neurosci. 2013; 13:80–93. [PubMed: 23224782]

58. Björnsdotter M, et al. Feeling good: on the role of C fiber mediated touch in interoception. Exp Brain Res. 2010; 207:149–155. [PubMed: 20963582]

59. Olausson H, et al. Unmyelinated tactile afferents signal touch and project to insular cortex. Nat Neurosci. 2002; 5:900–904. [PubMed: 12145636]

60. Morrison I. ALE meta-analysis reveals dissociable networks for affective and discriminative aspects of touch. Hum Brain Mapp. 2016; 37:1308–1320. [PubMed: 26873519]

61. Bennett RH, et al. fNIRS detects temporal lobe response to affective touch. Soc Cogn Affect Neurosci. 2014; 9:470–476. [PubMed: 23327935]

62. Kaiser MD, et al. Brain Mechanisms for Processing Affective (and Nonaffective) Touch Are Atypical in Autism. Cereb Cortex. 2015; doi: 10.1093/cercor/bhv125

63. Ackerley R, et al. An fMRI study on cortical responses during active self-touch and passive touch from others. Front Behav Neurosci. 2012; 6:51. [PubMed: 22891054]

64. McGlone F, et al. Touching and feeling: differences in pleasant touch processing between glabrous and hairy skin in humans. Eur J Neurosci. 2012; 35:1782–1788. [PubMed: 22594914]

65. Voos AC, et al. Autistic traits are associated with diminished neural response to affective touch. Soc Cogn Affect Neurosci. 2013; 8:378–386. [PubMed: 22267520]

66. Croy I, et al. Interpersonal stroking touch is targeted to C tactile afferent activation. Behav Brain Res. 2016; 297:37–40. [PubMed: 26433145]

67. Case LK, et al. Encoding of Touch Intensity But Not Pleasantness in Human Primary Somatosensory Cortex. J Neurosci. 2016; 36:5850–5860. [PubMed: 27225773]

68. Gazzola V, et al. Primary somatosensory cortex discriminates affective significance in social touch. Proc Natl Acad Sci. 2012; 109:E1657–E1666. [PubMed: 22665808]

69. Hogendoorn H, et al. Self-touch modulates the somatosensory evoked P100. Exp Brain Res. 2015; 233:2845–2858. [PubMed: 26105753]

70. Deschrijver E, et al. The interaction between felt touch and tactile consequences of observed actions: an action-based somatosensory congruency paradigm. Soc Cogn Affect Neurosci. 2016; 11:1162–1172. [PubMed: 26152577]

71. Schirmer A, et al. Squeeze me, but don't tease me: Human and mechanical touch enhance visual attention and emotion discrimination. Soc Neurosci. 2011; 6:219–230. [PubMed: 20711939]

72. Belin P, et al. Thinking the voice: neural correlates of voice perception. Trends Cogn Sci. 2004; 8:129–135. [PubMed: 15301753]

73. King AJ, Nelken I. Unraveling the principles of auditory cortical processing: can we learn from the visual system? Nat Neurosci. 2009; 12:698–701. [PubMed: 19471268]

74. Schirmer A, et al. The Socio-Temporal Brain: Connecting People in Time. Trends Cogn Sci. 2016; 20:760–772. [PubMed: 27615804]

75. Klasen M, et al. Multisensory emotions: perception, combination and underlying neural processes. Rev Neurosci. 2012; 23:381–392. [PubMed: 23089604]

76. Kreifelts B, et al. Audiovisual integration of emotional signals in voice and face: An event-related fMRI study. NeuroImage. 2007; 37:1445–1456. [PubMed: 17659885]

77. Kim J, et al. Abstract Representations of Associated Emotions in the Human Brain. J Neurosci. 2015; 35:5655–5663. [PubMed: 25855179]

78. Pourtois G, et al. The time-course of intermodal binding between seeing and hearing affective information. Neuroreport. 2000; 11:1329–1333. [PubMed: 10817616]

79. Ghose D, et al. Multisensory Response Modulation in the Superficial Layers of the Superior Colliculus. J Neurosci. 2014; 34:4332–4344. [PubMed: 24647954]

80. Miller RL, et al. Relative Unisensory Strength and Timing Predict Their Multisensory Product. J Neurosci. 2015; 35:5213–5220. [PubMed: 25834047]

81. Cappe, C., et al. Cortical and Thalamic Pathways for Multisensory and Sensorimotor Interplay. In: Murray, MM., Wallace, MT., editors. The Neural Bases of Multisensory Processes. CRC Press/ Taylor & Francis; 2012.

82. Schirmer A, Kotz SA. Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. Trends Cogn Sci. 2006; 10:24–30. [PubMed: 16321562]

83. Escoffier N, et al. Emotional expressions in voice and music: same code, same effect? Hum Brain Mapp. 2013; 34:1796–1810. [PubMed: 22505222]

84. Tanaka A, et al. I feel your voice. Cultural differences in the multisensory perception of emotion. Psychol Sci. 2010; 21:1259–1262. [PubMed: 20713633]

85. Darwin, C. The expression of the emotions in man and animals. John Murray; 1872.

86. Susskind JM, et al. Expressing fear enhances sensory acquisition. Nat Neurosci. 2008; 11:843–850. [PubMed: 18552843]

87. Bachorowski JA, Owren MJ. Sounds of emotion: production and perception of affect-related vocal acoustics. Ann N Y Acad Sci. 2003; 1000:244–265. [PubMed: 14766635]

88. Mitchell RLC, Phillips LH. The overlapping relationship between emotion perception and theory of mind. Neuropsychologia. 2015; 70:1–10. [PubMed: 25687032]

89. Bruce V, Young A. Understanding face recognition. Br J Psychol. 1986; 77:305–327. [PubMed: 3756376]

90. Yang DYJ, et al. An integrative neural model of social perception, action observation, and theory of mind. Neurosci Biobehav Rev. 2015; 51:263–275. [PubMed: 25660957]

91. Pitcher D, et al. Transcranial Magnetic Stimulation Disrupts the Perception and Embodiment of Facial Expressions. J Neurosci. 2008; 28:8929–8933. [PubMed: 18768686]

92. Adolphs R, et al. Neural systems for recognition of emotional prosody: a 3-D lesion study. Emot Wash DC. 2002; 2:23–51.

93. Banissy MJ, et al. Suppressing sensorimotor activity modulates the discrimination of auditory emotions but not speaker identity. J Neurosci Off J Soc Neurosci. 2010; 30:13552–13557.

94. Firestone C, Scholl BJ. Cognition does not affect perception: Evaluating the evidence for "top-down" effects. Behav Brain Sci. 2015; doi: 10.1017/S0140525X15000965

95. Rudrauf D, et al. Rapid interactions between the ventral visual stream and emotion-related structures rely on a two-pathway architecture. J Neurosci Off J Soc Neurosci. 2008; 28:2793–2803.

96. Tse CY, et al. The functional role of the frontal cortex in pre-attentive auditory change detection. NeuroImage. 2013; 83:870–879. [PubMed: 23871868]

97. Parr L, et al. MaqFACS: A Muscle-Based Facial Movement Coding System for the Rhesus Macaque. Am J Phys Anthropol. 2010; 143:625–630. [PubMed: 20872742]

98. Schirmer A, et al. Humans Process Dog and Human Facial Affect in Similar Ways. PLoS ONE. 2013; 8:e74591. [PubMed: 24023954]

99. Bryant GA. Animal signals and emotion in music: coordinating affect across groups. Front Psychol. 2013:4. [PubMed: 23378839]

100. Morimoto Y, Fujita K. Capuchin monkeys (Cebus apella) modify their own behaviors according to a conspecific's emotional expressions. Primates J Primatol. 2011; 52:279–286.

101. Schehka S, Zimmermann E. Affect intensity in voice recognized by tree shrews (Tupaia belangeri). Emot Wash DC. 2012; 12:632–639.

102. Schirmer A, et al. Tactile stimulation reduces fear in fish. Front Behav Neurosci. 2013; 7:167. [PubMed: 24319415]

103. Ghazanfar AA, Eliades SJ. The neurobiology of primate vocal communication. Curr Opin Neurobiol. 2014; 28:128–135. [PubMed: 25062473]

104. Parr L. The evolution of face processing in primates. Philos Trans R Soc Lond B Biol Sci. 2011; 366:1764–1777. [PubMed: 21536559]

105. Barber ALA, et al. The Processing of Human Emotional Faces by Pet and Lab Dogs: Evidence for Lateralization and Experience Effects. PloS One. 2016; 11:e0152393. [PubMed: 27074009]

106. Andics A, et al. Neural mechanisms for lexical processing in dogs. Science. 2016; 353:1030–1032. [PubMed: 27576923]

107. Lindell AK. Continuities in emotion lateralization in human and non-human primates. Front Hum Neurosci. 2013; 7:464. [PubMed: 23964230]

108. Schirmer A, et al. Reach out to one and you reach out to many: Social touch affects third-party observers. Br J Psychol. 2015; 106:107–132. [PubMed: 24628391]

109. Bediou B, et al. In the eye of the beholder: Individual differences in reward-drive modulate early frontocentral ERPs to angry faces. Neuropsychologia. 2009; 47:825–834. [PubMed: 19135071]

110. Palagi E, Mancini G. Playing with the face: Playful facial "chattering" and signal modulation in a monkey species (Theropithecus gelada). J Comp Psychol. 2011; 125:11–21. [PubMed: 21244137]

111. Mesulam MM. From sensation to cognition. Brain. 1998; 121:1013–1052. [PubMed: 9648540]

Trends

- Facial expression and perception have long been the primary emphasis in research. However, there is a growing interest in other channels like the voice and touch.

- Facial, vocal, and tactile emotion processing have been explored with a range of techniques including behavioral judgements, EEG/ERP, fMRI contrast studies and muti-voxel pattern analyses.

- Results point to similarities (e.g., increased responses to social and emotional signals) as well as differences (e.g., differentiation of individual emotions versus encoding of affect) between communication channels.

- Channel similarities and differences enable holistic emotion recognition – a process that depends on multisensory integration during early, perceptual and later, conceptual stages.
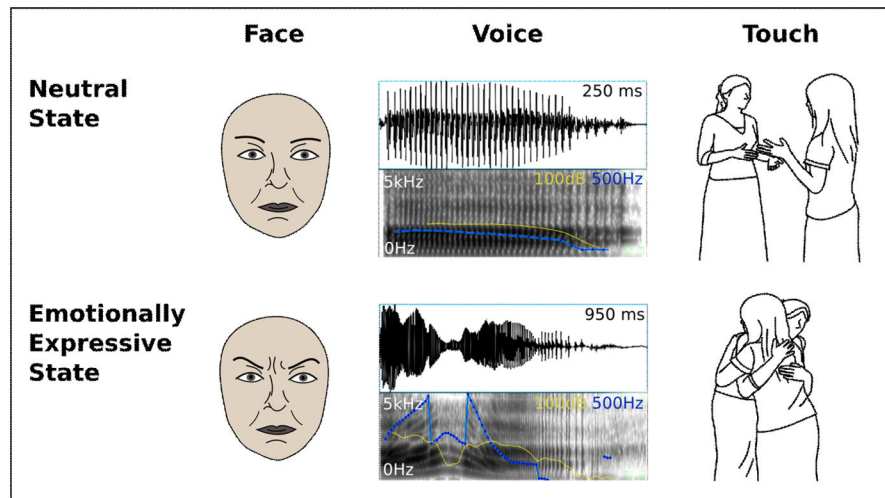
Figure 1.

Nonverbal emotion expression. Illustrated are the three modalities explored in this manuscript. (Left) The emotional facial expression differs from the neutral facial expression in action unit 4 (brow lowerer, corrugator supercilii, depressor supercilii), which forms part of anger, fear, and sadness displays. (Middle) The syllable "ah" spoken in a neutral and a happy tone are illustrated by oscillogram (top) and spectrogram (bottom). Vocalizing duration is longer, and pitch (blue) and intensity (yellow) are higher and more varied when the speaker is happy as compared to neutral. (Right) Compared to no-touch interactions, tactile interactions are perceived as more positive and arousing. Images taken from the Social Touch Picture Set [108].
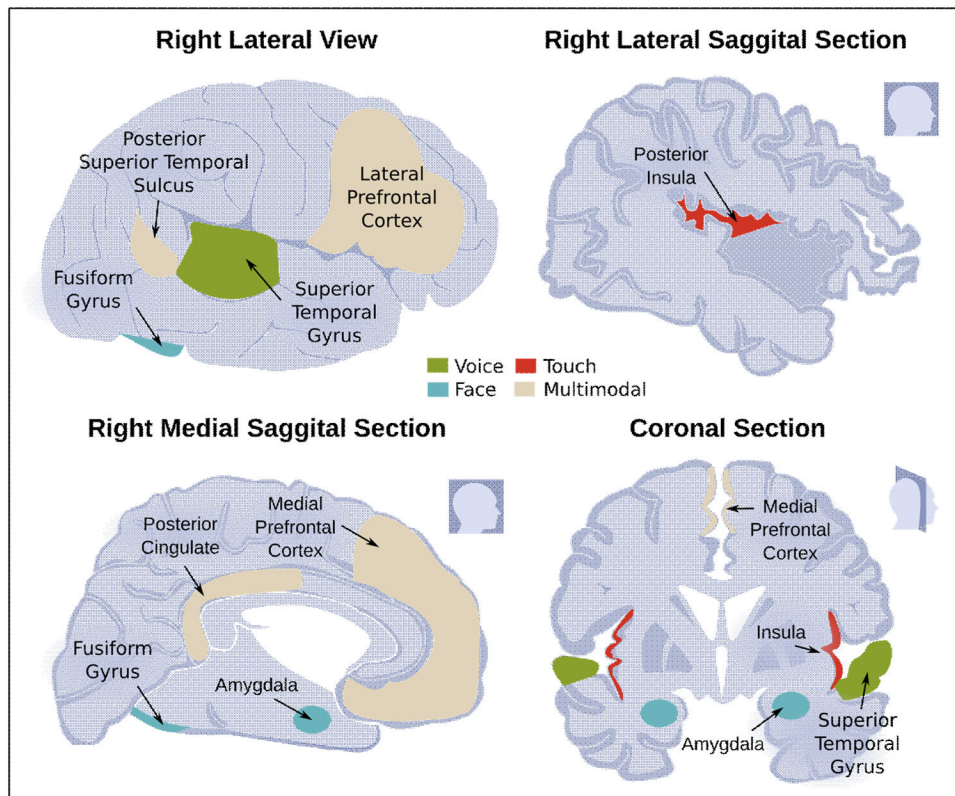
Figure 2.
Key brain regions involved in nonverbal emotion processing. Regions typically more active for emotional as compared to neutral stimuli are marked in green, blue, and red for voice, face, and touch, respectively. Regions typically more active for emotional multimodal as compared to unimodal stimulation are marked in beige.
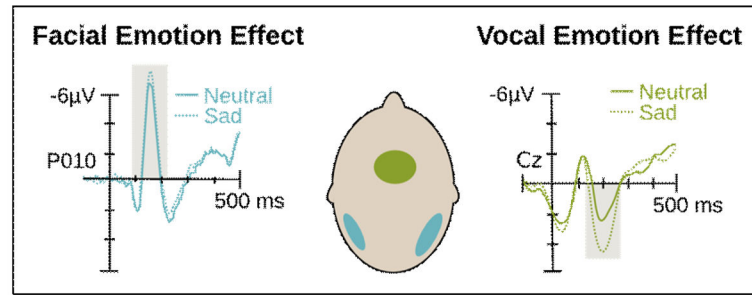
Figure 3.
ERP correlates of facial and vocal emotion processing. (Left) The facial emotion effect occurs for a negative deflection termed the N170 [109]. (Right) The vocal emotion effect occurs for a positive deflection termed the P200 [57]. Relevant components are marked by a gray box. (Middle) The scalp topography of the ERP peak maxima are illustrated on the cartoon head.
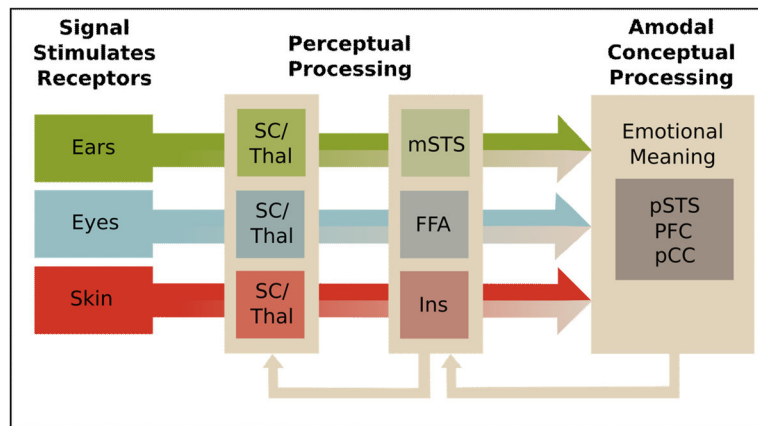
Figure 4.
Multimodal convergence of expressive signals in the brain. Beige color indicates the mixing of modalities which begins during early perceptual stages in the brainstem (Superior Colliculus, SC) and thalamus, and progresses at the level of the cerebrum, where physical cues (e.g., facial action units, vocal acoustics) are integrated and mapped onto existing emotion templates (i.e., emotional gestalt). At a late conceptual stage, individuals represent emotional meaning amodally. Higher-level representations can feed back and modulate lower-level representations.

Table 1

Comparing the processing characteristics of vision, audition, and touch for perceiving emotions.

Processing Characteristic	Face	Voice	Touch
Topographic maps	Yes	Yes	Yes
Slow and fast processing streams	Yes	Yes	Yes
Enhanced response to social vs non-social stimuli	Yes	Yes	Yes
Enhanced response to emotional vs. neutral expression	Yes	Yes	Yes
Peripheral receptor numbers	100M	16,000	~100,000
ERP markers within the first 200 ms	Yes	Yes	?
Perception requires temporal integration	No	Yes	Yes
Coordination with other modality	Voice	Face	?
Extent of processing before primary sensory cortex	Low	High	Low
Processing streams that bypass primary sensory cortex	Probably	Minimal?	Yes
Does stimulus perception induce feelings?	Sometimes	Yes	Yes