Emotion Recognition from Expressions in Voice and Face
– Behavioral and Endocrinological Evidence –

Dissertation

Zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades
"Doctor rerum naturalium"
der Georg-August-Universität Göttingen
im Promotionsprogramm Behavior and Cognition (BeCog)
der Georg-August University School of Science (GAUSS)

vorgelegt von

Adi Lausen

aus Diemrich, Rumänien

Göttingen, April 2019

Betreuungsausschuss

Prof. Dr. Annekathrin Schacht
Affektive Neurowissenschaft und Psychophysiologie, Institut für Psycholgie, Universität Göttingen

Prof. Dr. Lars Penke
Biologische Persönlichkeitspsychologie, Institut für Psycholgie, Universität Göttingen

Dr. Kurt Hammerschmidt
Kognitive Ethologie, Deutsches Primatenzentrum, Göttingen

Mitglieder der Prüfungskommision

Referent/in: Prof. Dr. Annekathrin Schacht
Affektive Neurowissenschaft und Psychophysiologie, Institut für Psycholgie, Universität Göttingen

Koreferent/in: Dr. Kurt Hammerschmidt
Kognitive Ethologie, Deutsches Primatenzentrum, Göttingen

Weitere Mitglieder der Prüfungskommision

Prof. Dr. Margarete Boos
Sozial- und Kommunikationspsychologie, Institut für Psycholgie, Universität Göttingen

Prof. Dr. Julia Fischer
Kognitive Ethologie, Deutsches Primatenzentrum, Göttingen

apl. Prof. Dr. York Hagmayer
Kognitionswissenschaft und Entscheidungspsychologie, Institut für Psycholgie, Universität Göttingen

Prof. Dr. Lars Penke
Biologische Persönlichkeitspsychologie, Institut für Psycholgie, Universität Göttingen

Tag der mündlichen Prüfung: 24. April 2019

Acknowledgments

This work could not have been carried out without the support of numerous people. Thus, I would like to take this opportunity to extend my deepest gratitude to those who have helped, supported and inspired me during this phase of my studies:

I am deeply grateful to my supervisor Prof. Dr. Annekathrin Schacht for the freedom she provided me to pursue a line of research that is still very unusual within the field of Affective Neuroscience. Thank you for encouraging my scientific adventures, for your support, comments, guidance, patience and encouragement throughout these three and a half years.

I am extremely grateful to my thesis-committee members, Dr. Kurt Hammerschmidt and Prof. Dr. Lars Penke, for their amicable support, helpful input and advice, as well as, for always being willing to meet me whenever I needed help.

Great gratitude and a big thank you to Christina Bröring, Dr. Wiebke Hammerschmidt, and Dr. Amanda Marshall, for their time, patience, help (and the list can go on) during the entire period of my PhD. I was really blessed having you not only as my strongest supporters, but for the greatest gift of all: your FRIENDSHIP.

I would like to thank Ronja Demel and Simon Stephan for their warm-heartedness, kindness and providing me with unquestioning support whenever I needed, as well as, to all members (past and present) of the ANaP Lab for the contributions they have made to my work. Special thanks to Dr. Annika Grass and Edmund Henniges for technical support and Isabel Nöthen, Saskia Brückner, Marc Köhler, Carlotta Dove, Annika Ziereis for helping in collecting high-quality data.

Thank you to my graduate RTG 2070/GEMI friends and colleagues which never failed to offer ingenious insights during our project discussions and for reminding me that there is life outside of graduate school. Though we carve new paths each day, may they always intersect from time to time.

Lastly, but certainly not least, I would like to thank Prof. Dr. Julia Fischer, Prof Dr. Margarete Boos and apl. Prof. Dr. York Hagmayer, for agreeing to be in my defense committee, Dr. Rebecca Jürgens for the administrative support and preparedness to help whenever needed, as well as, the two funding bodies supporting this project, Deutsche Forschungsgemeinschaft (DFG) and Leibniz-Science Campus Primate Cognition.

On a personal note, I would like to thank my partner for all the support. Words cannot express how grateful I am.

"Let's not forget that the little emotions are the great captains of our lives
and we obey them without realizing it"

Vincent van Gogh, 1889

Abstract

Emotion recognition is a key component of human social cognition and is considered vital for many domains of life. Studies measuring this ability have documented that performance accuracy in emotion recognition tasks is affected by various factors, ranging from gender, one's own confidence, hormonal fluctuations, to the modality of stimulus presentation (i.e., audio, visual). The majority of work has focused on the recognition of facial expressions. The results from the small amount of studies that made comparisons across the modalities of vocal and facial emotion recognition are contradictive, suggesting a lack of reliability across studies. Therefore, the main aim of this research project was to investigate the impact of above-mentioned factors on individuals' accuracy of performance while accounting for methodological shortcomings from previous research. Two independent but related studies were conducted. In **Study 1**, the first aim was to examine whether performance accuracy differs as a function of listeners' and speakers' gender. The second aim was to investigate the influence of vocal stimulus types and their related acoustic parameters on emotion recognition and confidence ratings. Additionally, it was explored whether the correct recognition of vocal emotions elicits confidence judgments. **Study 2** was preregistered and aimed to account for previous assumptions regarding males' 'poor' emotion recognition ability by investigating whether the modality of stimulus presentation (i.e., audio, visual, audio-visual) and hormonal fluctuations (i.e., testosterone, cortisol and their interaction) impact their performance accuracy and response time in emotion recognition tasks. In both studies, participants were asked to categorize the stimuli with respect to the expressed emotions in a fixed-choice response format. The results from **Study 1** showed that speakers' gender had a significant impact on how listeners' judged emotions from the voice, yet, no robust differences were observed regarding the performance accuracy of recognizing emotions by listeners' gender (*manuscript 1*). Additionally, the results obtained from this study replicate previous findings by showing that participants could recognize emotions based on differential acoustic patterning. They further add to previous research by demonstrating that emotional expressions are more accurately recognized and confidently judged from non-speech sounds than from emotionally inflected speech. Moreover, they showed that listeners who were better at recognizing vocal expressions of emotion were also more confident in their judgments (*manuscript 2*). The results from **Study 2** indicated that emotion recognition accuracy and response time are greatly improved for the audio-visual presentation of emotional expressions. In addition, they showed that happy expressions are identified faster and with greater accuracy from faces than voices, while angry expressions are better recognized in voices compared to faces. Finally, the overall effect sizes of testosterone by cortisol interaction on emotion recognition accuracy and response time were small yet significant (*manuscript 3*). The combined findings from both studies explain inconsistencies in the existing literature by highlighting the importance of distinguishing between these factors when assessing emotion recognition ability. This research project actively contributes to a scientific domain that is currently re-writing our

understanding on the role these factors play for the recognition of emotions. It hereby paves the way for impactful future research.

Keywords: emotion recognition, accuracy, basic emotions, prosody, vocal bursts, facial expressions, gender differences, acoustic parameters, reaction time, confidence judgements, testosterone, cortisol, dual-hormone hypothesis

Contents

Chapter 1

General Introduction

Conducting research on emotion is at once a fascinating and difficult endeavor. It is fascinating because emotions are of central importance to our daily life, and it is difficult because even though there has been extensive research on emotions, there is not yet one unified and generally accepted theory on emotion. Rather, various different approaches to and multiple aspects of emotions have been studied.

1.1 A brief primer on emotion

> "Everyone knows what an emotion is, until asked to give a definition. Then, it seems, no one knows" (Fehr & Russell, 1984, p.464)

As there are no generally agreed on criteria for what should count as an 'emotion' and what should not, this concept is hard to define (Frijda, Scherer, & Sander, 2009; Scarantino & de Sousa, 2018). Although there are different theories to explain what emotions are and how they operate (see Izard, 2010; K. Scherer, 2009, for details), most researchers agree that emotions are relatively brief, intense reactions which serve a coordinating role by automatically triggering a set of concomitant responses (i.e., physiology, behavior, experience, communication). These enable individuals to deal quickly with problems or opportunities in their external or internal environment (e.g., Juslin & Laukka, 2003; Keltner & Gross, 1999; K. R. Scherer & Moors, 2019). There is also a consensus among researchers that emotions consist of several components: *cognitive appraisals* (e.g., you appraise the situation as dangerous), *subjective feelings* (e.g., you feel afraid), *hormonal and physiological responses* (e.g., stress hormones are released and your heart starts to beat faster), *vocal, facial,* and *bodily expressions* (e.g., you scream), *action tendencies* (e.g., you run away) and *regulation* (e.g., you try to calm yourself) (e.g., Planalp, 1999; K. R. Scherer & Moors, 2019; Shuman & Scherer, 2015). However, there is disagreement about how emotions should be modeled or conceptualized: as discrete categories (e.g., Ekman, 1992), dimensions (e.g., Russell, 1980), prototypes (e.g., Shaver, Schwartz, Kirson, & O'connor, 1987), or component processes (e.g., K. R. Scherer, Banse, & Wallbott, 2001; K. R. Scherer et al., 2000). Proponents that favor a discrete approach towards modelling emotions suggest that there is an innate basic set which is given to us by nature (Darwin, 1872; Izard, 2007).

This set of innate emotions has been selected for their adaptive value. In consequence, the automatic mechanisms they trigger are capable of regulating interactions with the proximal environment. At the same time, they provide effective responses, both instrumental and communicative, in relation to the relevant situation for survival (Levenson, 2011; Shariff & Tracy, 2011b; Tooby & Cosmides, 2008). In contrast, theorists that favor a dimensional approach to model emotions argue that emotions are products of nurture rather than nature. They suggest that emotions are socially constructed and that language, culture, conceptual knowledge, as well as contextual factors shape our emotional responses along continuums such as valence (negative to positive) and arousal (calm to excited) (e.g., Barrett, Lindquist, & Gendron, 2007; Barrett, Mesquita, & Gendron, 2011; Kuppens, Tuerlinckx, Russell, & Barrett, 2013; Lindquist, 2013). Componential emotion theorists combine elements of dimensional models (i.e., emotions as emergent results of underlying dimensions) with elements of discrete theories (i.e., emotions have different subjective qualities). They postulate that the experience of an emotion is determined by a series of cognitive evaluations on different levels of processing (e.g., Ellsworth & Scherer, 2001; K. R. Scherer, 2009; Shuman & Scherer, 2014) that account for both individual- and cultural differences (Mortillaro, Meuleman, & Scherer, 2012).

Beyond these theoretical debates, the predictions of each approach have drawn support from behavioral, neuropsychological, psychophysiological and neuroimaging studies (e.g., Bestelmeyer, Kotz, & Belin, 2017; Damasio et al., 2000; Murphy, Nimmo-Smith, & Lawrence, 2003; K. R. Scherer & Ellgring, 2007; Vytal & Hamann, 2010; Wyczesany & Ligeza, 2015), as well as, from findings on analogous or homologous responses in non-human primates and other mammals (e.g., J. Fischer, Metz, Cheney, & Seyfarth, 2001; Panksepp, 2007; Parr, Cohen, & De Waal, 2005; Parr, Waller, & Fugate, 2005). Thus, one needs to emphasize that these contrasting approaches to model emotions, are not necessarily mutually exclusive, but rather offer different descriptions of the same underlying phenomenon while highlighting different aspects in the emotion-generation process (Harmon-Jones, Harmon-Jones, & Summerell, 2017). It is important, however, to be clear whether we are testing hypotheses about emotions (commonly viewed as distinct categories or discrete entities), affect (generally conceptualized as two continuous but bounded dimensions) or some combination of the two [e.g., when we ask whether (discrete) emotions or (dimensional degrees of) affect capture more variance in a given situation]. Depending on the area of inquiry, we may get different answers that speak for or against any particular approach. For instance, Mortillaro et al. (2012), suggested that classifying basic emotion expressions as discrete categories may be a fruitful approach for studying emotion recognition. Conversely, when the goal is to detect broad elements like valence or arousal a dimensional model may provide more satisfactory results. When the aim is to address a potentially large number of affective states, a component model may offer more advantages over the discrete and dimensional frameworks of emotion. In line with this suggestion and with previous findings highlighting that emotional expressions are perceptually coded in terms of their conformity to prototype expressions that correspond to basic emotions (e.g., Calder, Young, Perrett, Etcoff, & Rowland, 1996; de Gelder, Teunisse, & Benson, 1997;

Laukka, 2005), the present research implemented a categorical approach in its endeavor to examine various factors that impact on individuals' ability to recognize basic emotions.

1.2 Emotion expressions: signals for emotion recognition

Emotion expressions were argued to be the "grammar of social interaction" as they structure how individuals relate to one another (Keltner, Tracy, Sauter, Cordaro, & McNeil, 2016). They play a crucial role in many social processes and serve as a window into reactions, intentions, and likely future behaviors through two interrelated mechanisms (Laukka, 2008). First, by expressing emotions we can communicate important information to others and thereby influence their behaviors or attitudes. Second, the recognition of others' emotional expressions allows us to make quick inferences about their internal states or intentions (Côté & Hideg, 2011; Juslin, 2013). Studies on the social purpose of emotional expressions suggested that they originally evolved to serve internal, physiological functions, and later came to serve more social communicative purposes (Shariff & Tracy, 2011b). This shift is thought to have occurred through a process of ritualization (see De Waal, 2003; Parr, 2003, for details), wherein the nonverbal behaviors occurring with particular emotions (e.g. eyes widening with fear) became reliably associated with those emotions, and, as a result, came to serve as a signal of them (Shariff & Tracy, 2011b). Therefore, emotion expressions became exaggerated into the highly recognizable and prototypical forms we observe them in today, which function to signal important information to observers (Parr, 2003; Shariff & Tracy, 2011b). As signals they have a major impact on social communication and, therefore, it has been argued that their fast and accurate recognition serves critical adaptive advantages (Shariff & Tracy, 2011b) [see also (Barrett et al., 2011; Shariff & Tracy, 2011a, for a hot debate)|. Although emotion expressions are multimodal patterns of behavior involving bodily movement, gaze, gestures, touch and even scents (see Keltner et al., 2016, for a review on the signaling properties of these expressions) the emphasis of this introduction will be on vocal and facial expressions as they are of central relevance for the present research.

1.2.1 Emotion expressions

Vocal expressions

To vocalize (emotions), our lungs have to produce energy by filling the trachea below the closed glottal folds with air. Together with motor commands to the laryngeal musculature, this subglottal air pressure brings about phonation (i.e., vibration of vocal folds release air pulses into the supraglottal vocal tract). In order to articulate, the series of pulses which are released in the supraglottal vocal tract are varied by tongue, lips, or jaw movements (see Kappas, Hess, & Scherer, 1991; K. R. Scherer, 1986, for an overview of the voice production system and its major determinants). In his review of the physiological and neurological systems that control the production mechanisms of vocalization, K. R. Scherer (1989) argued that linguistic speech production is primarily controlled by the neocortex, whereas the emotional vocalization is controlled by the limbic system. He proposes that

effects of emotional arousal on the speech production process are primarily produced by tonic activation of the autonomic and somatic nervous system. Their effects, such as respiration, phonation, and articulation influence the nature of the vocal output.

In research on human speech, one needs to distinguish between segmental aspects, which carry linguistic information (i.e., lexical, syntactic and semantic), and suprasegmental aspects, which carry a mixture of paralinguistic and non-linguistic information (Johnstone, Van Reekum, & Scherer, 2001; Murray & Arnott, 1993). The non-linguistic information carried by the voice includes indicators of the speaker's age, gender, state of health, their cultural and educational background (Kappas et al., 1991) and, of central interest here, their emotional state. Pitch, loudness, duration, together with speech rate, voice quality (and others not listed here) combine to form the paralinguistic attributes of speech or prosody (see ***Figure 1***). These perceptual properties and their primary acoustic coun-

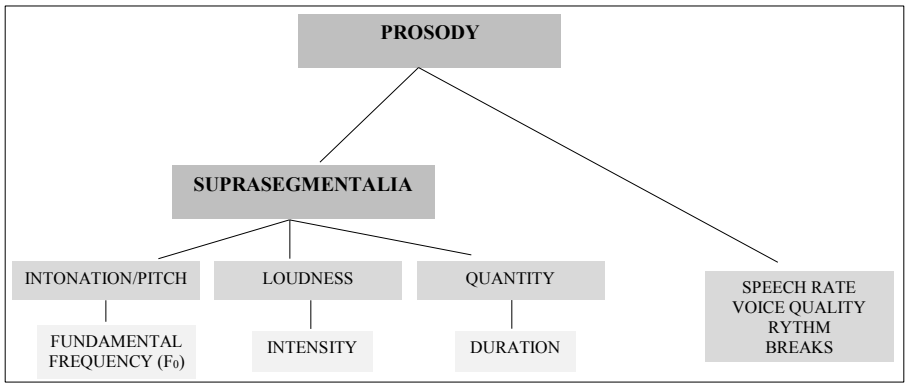
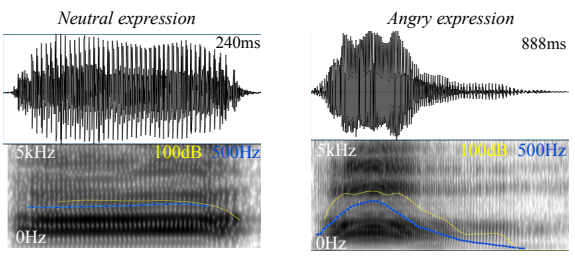


Figure 1 | Suprasegmental parameters of emotional speech (prosody). Figure adapted from Moebius (1993, p.9).

terparts (see *manuscript 2*, for details) were found to play a major role for the expression of emotions (e.g., Juslin & Laukka, 2003; Laukka et al., 2016; Patel, Scherer, Björkner, & Sundberg, 2011; Paulmann, Pell, & Kotz, 2008; W. F. Thompson & Balkwill, 2009), with research identifying specific vocal expression patterns for each basic emotion (see ***Table 1***, for an overview). Varying these vocal attributes during speech generation represents one way for humans to convey internal states and emotions, hereafter referred to as prosody.[1]

Table 1 | Acoustic profiles of vocal emotions

Emotion	*Vocal expression parameters*
Angry	Pitch and intensity increase, speech rate faster, breathy and tense tone
Disgust	Pitch decrease or increase, intensity decrease, slower speech rate, grumble chest tone
Fear	Pitch and intensity increase, speech rate faster, scratchy, breathy and irregular tone
Happiness	Pitch and intensity increase, speech rate faster or slower, breathy and blaring tone
Sadness	Pitch and intensity decrease, speech rate slower, lax and resonant tone
Surprise	Pitch increase, speech rate faster, breathy tone



Note: Illustrated are two examples of stimuli explored in the present research. The vocal burst 'ah' uttered in a neutral and an angry tone is illustrated on an oscillogram (top) and a spectrogram (bottom). Vocalizing duration is longer, pitch (blue) and intensity (yellow) are higher and more varied when the speaker is angry compared with neutral. The sounds are from *Montreal Affective Voices* (Belin et al., 2008). The summary on emotion-specific voice patterns is based on the results reported in a meta-analysis (e.g., Juslin & Laukka, 2003) and single studies (Hammerschmidt & Jürgens, 2007; Schuller, Rigoll, & Lang, 2004; Cowie et al., 2001; Johnstone & Scherer, 2000; Trainor, Austin & Desjardins, 2000; Banse & Scherer, 1996).

[1] In addition to its expressive function (i.e., to convey internal states and emotions), prosody can also be used as a representative speech act (i.e., to provide information) or a directive act (i.e., to ask for something). Accordingly, it has been argued that prosody has both, an emotional and linguistic function (see Paulmann, Titone, & Pell, 2012; Shih & Kochanski, 2002, for details).

One other way for humans to express vocal emotions is with vocal/affect bursts, defined as brief non-linguistic sounds (Hawk, Van Kleef, Fischer, & Van der Schalk, 2009) that occur in between speech incidents or in the absence of speech (Cowen, Elfenbein, Laukka, & Keltner, 2018). Examples include cries, laughs, screams, growls, moans, babbling, ahhs and oohs (Belin, Fillion-Bilodeau, & Gosselin, 2008; K. R. Scherer, 1994). Vocal bursts are considered precursors of speech (Cordaro, Keltner, Tshering, Wangchuk, & Flynn, 2016) and are thought to parallel animal vocalizations (Krumhuber & Scherer, 2011). Research has shown that primates emit vocalizations in terms of emotional urges (J. Fischer & Price, 2017) that are specific (goal-directed) to predators, food, affiliation, care, sex, and aggression (e.g., K. Hammerschmidt & Fischer, 2019; Snowdon, 2003). Similar to non-human primates' alarm calls, screaming is arguably one of the most relevant communication signals for survival in humans, providing a behavioral advantage by increasing speed and accuracy of spatially localizing a potential threat in environment (Arnal, Flinker, Kleinschmidt, Giraud, & Poeppel, 2015). For instance, brain imaging studies showed that the amygdala and the interconnected limbic regions are activated in parents' listening to infants' cries compared to non-parents (Seifritz et al., 2003). Further, the periaqueductal gray, a region of the midbrain believed to promote protective caregiving responses in adults, is involved in the processing of both infant cries and laughs (Parsons, Young, Joensson, et al., 2014). Neuroendocrinal studies tell a similar story, showing heightened testosterone in fathers' responses to infant cries compared to non-fathers (Fleming, Corter, Stallings, & Steiner, 2002) and cortisol elevations in adult- but not teen mothers' (Giardino, Gonzalez, Steiner, & Fleming, 2008).

As they convey information about features of the environment which orients hearers' actions, it has been argued that vocal bursts are more than just fleeting ways through which we communicate emotion (Cowen et al., 2018). Together with prosody they provide important information relevant to perceivers, which is useful in guiding subsequent behavior.[2]

Facial expressions

Similar to voices, facial expressions of emotion are known to be highly relevant for social communication. Research has identified vital information that can be inferred from these expressive actions (see Jack & Schyns, 2015, for details) and demonstrated their importance for social interactions (see Balconi, 2010; Keltner et al., 2016, for reviews). For adaptive reasons and because of their ubiquity, facial expressions were argued to be signals of high biological and social value (K. L. Schmidt & Cohn, 2001; Smith & Rossit, 2018). Just like

[2]The importance of prosody and vocal bursts is apparent in domains of social interaction which transcend the boundaries of this work (i.e., other than emotion). For instance, it has been found that solely based on the tone of our voice people can infer other individuals' intentions (e.g., Hellbernd & Sammler, 2016), confidence (e.g., Jiang & Pell, 2015, 2017), attractiveness (e.g., Fraccaro et al., 2013; Xu, Lee, Wu, Liu, & Birkholz, 2013), social class (e.g., Gregory Jr & Webster, 1996; Kraus, Park, & Tan, 2017), dominance (e.g., Hodges-Simeon, Gaulin, & Puts, 2010) and trustworthiness (e.g., Ponsot, Burred, Belin, & Aucouturier, 2018). In addition, research has shown that from laughs, adults can infer a person's rank within a social hierarchy (Oveis, Spectre, Smith, Liu, & Keltner, 2016) or whether two individuals are friends or strangers (Bryant et al., 2016; Smoski & Bachorowski, 2003).

acoustical features within the voice, a combination of action units (i.e., physical features based on muscle movement of the face) are believed to give a reliable impression of the underlying emotions displayed in the face (Schirmer & Adolphs, 2017, see ***Table 2***, for an overview). Facial expressions of emotion arise from the existence of two neural pathways of

Table 2 | Physical features of facial expressions

Emotion	*Facial expression action units*
Angry	Eyebrows are drawn together and pulled down, upper and lower eyelids are pulled up, lips are tightened or with an open mouth
Disgust	Eyebrows pulled down, nose wrinkled, upper lip pulled up, lips loose
Fear	Eyebrows are raised, eyes are wide open, lips are usually stretched and tense and the mouth may be open or closed
Happiness	Lower eyelids are tense, lip corners are pulled up raising the cheeks, wrinkles around the eye
Sadness	Inner corners of eyebrows raised, eyelids loose, lip corners pulled down
Surprise	Eyebrows are raised, eyes are wide open, mouth hangs open and pupils are dilated

Note: Illustrated are two examples of stimuli explored in the present research. Different emotional expressions can be characterized by combinations of action units (basic sets of muscle movements) whereby a single unit can participate in multiple emotional expression. 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 (see Matsumoto, Keltner, Shiota, O'Sullivan & Frank, 2008; Ekman & Friesen, 1976, for details). The facial displays are from *Radboud Faces Database* (Langner et al., 2010).

innervation of the facial musculature, each one originating in a different area of the brain (Matsumoto & Lee, 1993; Müri, 2016). Involuntary, spontaneous emotional expressions are activated by neurons in the subcortical areas of the brain, while the voluntary, deliberate facial expressions by the frontal regions and the motor cortex (Cacioppo, 2013; Hwang & Matsumoto, 2016). The importance of these two pathways for the production of facial expressions of emotion found support in clinical reports of brain-damaged patients, as well as, in neuroimaging studies (see George, 2013; Posamentier & Abdi, 2003, for comprehensive reviews). In addition, neurophysiological research has demonstrated that emotional information from faces is detected rapidly 100 ms after stimulus onset, and different facial expressions are discriminated within an additional 100 ms with higher amplitudes for emotional compared to neutral expressions (e.g., Bublatzky, Gerdes, White, Riemer, & Alpers, 2014; Eimer & Holmes, 2007; W. Hammerschmidt, Kulke, Broering, & Schacht, 2018; W. Hammerschmidt, Sennhenn-Reulen, & Schacht, 2017; Hinojosa, Mercado, & Carretié, 2015).

Given the importance of emotional expressions for social interaction (K. R. Scherer, Clark-Polner, & Mortillaro, 2011) the face has gained particular attention among researchers and contains an impressive record of empirical research (i.e., from developmental to comparative studies). A full review of the literature is far beyond the scope of the current work. It is relevant, however, to highlight certain aspects on the communicative role of facial expressions in social interaction. First, facial expressions can rapidly convey information about the emotional state of others, their intentions, attitudes and likely future behaviors (Jang & Elfenbein, 2015). Second, they evoke complementary emotions (e.g., anger may trigger fear), contagion and regulation effects in others (see Van Kleef, 2016, for a comprehensive review). Third, they serve as incentives or disincentives for other peoples' behavior, as evidenced by the strength of emotional rewards and punishments in learning and shaping behavior (Keltner, Kring, & Bonanno, 1999).

The minimal review above shows that facial expressions of emotion play a pivotal role

in regulating social interaction. If we accept the assumption that emotional expressions have evolved in part because of their communicative functionality, it stands to reason that facial-, vocal and other non-verbal displays (posture, gesture) equally shape our social interactions (Van Kleef, 2017). However, this might be the case only under conditions where emotional expressions can be accurately perceived and it may also largely depend on individuals' emotion recognition ability.

1.2.2 Emotion recognition

As terminology is not always used in a consistent manner in the empirical literature, it is not easy to determine researchers' theoretical assumptions about whether they measure emotion recognition or other socio-cognitive processes such as theory of mind (ToM) or empathy. These terms are often used interchangeably due to the fact that they are so interrelated (see Batson, 2009; Cuff, Brown, Taylor, & Howat, 2016; Mitchell & Phillips, 2015; Olderbak & Wilhelm, 2017, for a review on used definitions). From the perspective espoused in the present research, emotion recognition is defined as the most fundamental and basic component of social cognition (e.g., Arioli, Crespi, & Canessa, 2018; Frith & Frith, 2007), emotional intelligence (e.g., Mayer, Salovey, Caruso, & Sitarenios, 2001; Roberts, MacCann, Matthews, & Zeidner, 2010) and emotional competence (e.g., Bänziger, Grandjean, & Scherer, 2009). Independent of whether researchers conceive emotion recognition as a part of social cognition, emotional competence or emotional intelligence, they all consider this ability a crucial prerequisite of social interactions that allows individuals of the same species (conspecifics) to make sense of others behavior patterns (Arioli et al., 2018; Schlegel, Fontaine, & Scherer, 2017). In human communication, the accurate recognition of emotional expressions was argued to be of paramount importance for social interactions and personal relationships (e.g., A. H. Fischer & Manstead, 2008). Difficulties in correctly recognizing these signals can lead to problematic social relationships and eventually result in the development and maintenance of psychopathology (e.g., Goldman & Sripada, 2005; Keltner & Kring, 1998; C. G. Kohler, Walker, Martin, Healey, & Moberg, 2009; Marsh & Blair, 2008).

Results from numerous studies and meta-analyses consistently reported that in healthy participants the accuracy of recognizing basic- and other emotional displays (e.g., awe, amusement, shame) from the voice and face is above chance levels (Cordaro et al., 2018; Elfenbein & Ambady, 2002b; Laukka et al., 2016), with scores approaching 70% accuracy for prosodic utterances (e.g., Juslin & Laukka, 2003; Paulmann & Uskul, 2014; Pell, Paulmann, Dara, Alasseri, & Kotz, 2009), 79% for vocal bursts (Cordaro et al., 2016) and above 70% for static and dynamic facial expressions (K. R. Scherer & Scherer, 2011). There is, however, considerable individual variability in the extent to which people are able to recognize target emotions from prototypical displays (Israelashvili, Oosterwijk, Sauter, & Fischer, 2019; Keltner, Sauter, Tracy, & Cowen, 2019). Several factors have been proposed to account for individuals' differential success in emotion recognition from nonverbal expressions. These are briefly summarized below.

Biological sex

One of the most discussed factors that may influence performance accuracy in emotion recognition tasks is biological sex (Kret & De Gelder, 2012). Despite the widely-held assertion of female 'superiority' in emotion recognition tasks, a closer examination of the literature reveals a more mixed picture. While some findings indicate that females are overall (i.e., across all emotions) better than males at recognizing emotions (e.g., Hall, 1984; McClure, 2000; Wingenbach, Ashwin, & Brosnan, 2018), other studies report only a small effect [$d = 0.19$] (e.g., A. E. Thompson & Voyer, 2014) or no sex differences (e.g., Lyusin & Ovsyannikova, 2016; Matsumoto & Hwang, 2011; Rahman, Wilson, & Abrahams, 2004). Inconsistent patterns are also observed when examining sex differences for specific emotion categories or with regard to the performance accuracy at categorizing certain emotions when expressed by female or male actors. For instance, an earlier study examining sex differences in the decoding (i.e., recognition) and encoding (i.e., expression) of negative emotions (i.e., anger, disgust, fear and sadness) showed that females exceeded males in their ability to recognize these emotions independent of expresser (actor) gender, except for anger where males were superior to females in recognizing this emotion when expressed by a male actor (Rotter & Rotter, 1988). Some subsequent studies tell a similar story by showing that males' angry displays are detected significantly more rapidly by male than female observers (e.g., M. A. Williams & Mattingley, 2006). However, other studies tell a different story by reporting that, independent of participants' sex, performance accuracy is higher for portrayals of anger and fear when expressed by male actors (e.g., Bonebright, Thompson, & Leger, 1996) whereas fear and disgust are better identified when expressed by female actors' (e.g., Collignon et al., 2010). The picture is further complicated by some studies showing that fear, disgust, happiness and sadness are better recognized by females (e.g., Hall & Matsumoto, 2004; Lee et al., 2013), while other studies fail to show this pattern, with the exception of disgust (e.g., Connolly, Lefevre, Young, & Lewis, 2019).

Although explanations for these sex-based behavior patterns in recognition accuracy range from socio-cultural influences to psychological and biological dispositions (see Babchuk, Hames, & Thompson, 1985; Chaplin, 2015; Davis et al., 2012; A. H. Fischer & LaFrance, 2015; Hall, 1984; Hyde, 2014; Schirmer, 2013), only limited consensus can be reached given the nature of the existent literature. Furthermore, the literature is almost entirely reliant on studies of facial expression recognition, while for other expressive domains such as the voice or body postures, only little is known about these sex-based behavior patterns. This paucity makes it even more difficult to establish a reliable generalization of the findings.

Hormones

Steroid hormones (i.e., estradiol, progesterone, testosterone) were highlighted as an additional predictor that might impact on individuals' emotion recognition ability (e.g., Gignell, Hornung, & Derntl, 2019; Van Honk & JLG Schutter, 2007). While some studies indicate that in naturally cycling women high levels of estradiol and progesterone negatively correlate with performance accuracy and reaction times in emotion recognition tasks (e.g., Derntl, Kryspin-Exner, Fernbach, Moser, & Habel, 2008; Derntl, Windischberger, et al.,

2008; Kamboj, Krol, & Curran, 2015), little is known about the effects of testosterone or other hormones, such as cortisol, on emotion recognition. This is surprising since steroid- and stress related hormones (i.e., cortisol) are highly expressed in the limbic system and the hypothalamus, both areas associated with the processing of emotions (see Gignell et al., 2019; Hakamata et al., 2017, for details). Concerning the effects of testosterone on emotion recognition, the few existing studies yield rather contradictory results. For instance, in a placebo controlled cross-over study, Van Honk and JLG Schutter (2007) examined whether females' ability to recognize 'threat'- (anger, disgust, fear) vs. 'non-threat' (happiness, sadness, surprise) facial expressions differs after testosterone administration. The authors reported a significant decrease in performance accuracy for threat related expressions, however, when the emotion categories were separately analyzed this effect remained significant only for angry faces. A similar pattern was found by Rukavina et al. (2018), who reported that with increasing levels of testosterone males' emotion recognition accuracy decreases. However, other researchers found either a positive association between testosterone levels and males' emotion performance accuracy (Vongas & Al Hajj, 2017) or no association at all (Derntl et al., 2009). Inconclusive results were also found for the influence of cortisol on emotion recognition. While Feeney, Gaffney, and O'Mara (2012) found that higher levels of cortisol are associated with quicker identification of emotional facial expressions, as well as, with a greater tendency to judge neutral faces as emotional, Duesenberg et al. (2016) found no support that increases in cortisol levels influence emotion recognition in healthy young individuals.

Taken together, the reported results provide tentative evidence that testosterone or cortisol alone influence performance accuracy or reaction time in explicit emotion recognition tasks. Most of the above-mentioned studies have small sample sizes ($N < 85$) and, as such, may be underpowered to detect what are unlikely to be large effect sizes. Furthermore, these studies are hard to compare due to the heterogeneity of paradigms and tasks used (implicit or explicit emotion recognition tasks). Also, the studies have large discrepancies with respect to the collection of hormone samples (i.e., blood or saliva) or their storage (-20°C or -80°C). The lack of studies, as well as, methodological differences and heterogeneity of paradigms make it very difficult to draw clear conclusions about the role of steroid- or stress related hormones on emotion recognition.

Further related factors

Both the vocal and facial emotion recognition literature has explored the relationship between different personality traits, empathy, mood, confidence judgments and emotion recognition accuracy (although, as with hormones and sex differences literature, far more emphasis has been put on detecting emotions from faces). For instance, in the vocal emotion literature, extraversion and conscientiousness have been associated with better recognition of emotions from the voice, but only in males (Burton et al., 2013). In contrast, Terracciano, Merritt, Zonderman, and Evans (2003) found a positive relationship between the recognition of vocal emotions and openness to experience. Similarly, in the facial emotion literature, some studies have found a link between accuracy of performance and

openness to experience and conscientiousness (e.g., Matsumoto et al., 2000). Other studies have emphasized the importance of extraversion and neuroticism. While some researchers have argued that extraverted individuals perform better on facial emotion recognition tasks (e.g., Matsumoto et al., 2000; K. R. Scherer et al., 2011), other studies have failed to evidence this relationship (e.g., Cunningham, 1977). Similarly, neuroticism has been linked to both lower (Matsumoto et al., 2000) and higher (Cunningham, 1977) performance accuracy in emotion recognition tasks. It is, thus, apparent that the relationships between personality traits and emotion recognition are not wholly consistent. Similar (i.e., inconsistent) patterns are reported across studies for the correspondence between empathy, mood and emotion recognition accuracy (see Altrov, Pajupuu, & Pajupuu, 2013; Clore, Schwarz, & Conway, 1994; Gaddy & Ingram, 2014; Gery, Miljkovitch, Berthoz, & Soussignan, 2009; Matt, Vázquez, & Campbell, 1992; Schmid & Mast, 2010; Zaki, Bolger, & Ochsner, 2008, for studies on empathy and mood). However, for confidence judgements, too few studies have been performed to gain traction on possible effects on emotion recognition accuracy. For instance, Kelly and Metcalfe (2011) found that emotional expressions that are confidently understood promote correct and confident interpretations in emotion recognition tasks. Insofar, this is the only study that specifically examined the impact confidence judgments have on peoples' ability to recognize emotions.

Finally, another factor that has been repeatedly suggested to impact on individuals' ability to recognize emotions is the modality of stimulus presentation. The audio-visual integration of expressive stimuli has a long history of producing additive effects in emotion recognition accuracy (e.g., De Gelder & Vroomen, 2000; Massaro & Egan, 1996; Paulmann & Pell, 2011; Vroomen, Driver, & De Gelder, 2001). Across these studies, there is agreement that combining complementary stimuli make the expression appear more intense, thus, increasing the likelihood of an emotion to be decoded correctly. Despite a growth in the number of studies showing that the audio-visual presentation of emotional stimuli increases performance accuracy across and for specific emotion categories (e.g., Bänziger et al., 2009; Collignon et al., 2008; Kreifelts, Ethofer, Grodd, Erb, & Wildgruber, 2007), the comparison of results for the audio- and the visual presentation of emotional stimuli are often contradictory (Elfenbein & Ambady, 2002b; Kraus, 2017; Lambrecht, Kreifelts, & Wildgruber, 2014; Waaramaa, 2017). For instance, an early meta-analysis reports that happiness is the most accurately recognized emotion in the face, while anger is the best recognized emotion in the voice (Elfenbein & Ambady, 2002b). However, more studies are needed with the direct aim of assessing the association between an emotion and a particular modality of expression. Without these targeted comparisons, results to date may reflect either a real preference over the other modality (indicating that these modalities do not merely carry redundant information but rather each may have certain specialized functions in the communication of emotion), or the apparent pattern could be a meaningless artifact of the types of emotions included in previous studies assessing single-channel communication.

Taken altogether, the results from the various studies mentioned-above do provide a wealth of information regarding potential factors that might impact on individuals' ability

to recognize emotions. As pointed out the reported results are often contradictory, and, therefore, our understanding with regard to emotion recognition is far from complete and warrants further investigation.

1.3 Present research approach and aims

Previous research has found that in comparison to naturally occurring expressions, simulated (i.e., play-acted) emotions are not necessarily expressed in a more exaggerated or stereotypical fashion (e.g., Jürgens, Hammerschmidt, & Fischer, 2011). Based on this evidence, the present research made use of play-acted expressions of emotion in its endeavor to assess the recognition of vocal and facial expressions of emotion. However, the use of actor portrayals has been criticized due to concerns about ecological-validity as they have been characterized as "caricatures" of real-life emotional expressions (Barrett, 2006). Therefore, their use in the present research requires a short clarification. The aim of this research was to examine the prototypical representation of basic emotions in social communication and not to study the nature of spontaneous emotion expressions in real life. The emotion recognition tasks in the present research required participants to judge what emotion is *represented* by the portrayal and not what emotion the actor feels. Studying an expressive code requires a careful selection of portrayed expressions that are understood by the decoder. The present investigation consisted of two independent but related studies.

Study 1: Recognizing emotions from the voice

General information: The majority of researchers who have studied the recognition of vocal expressions of emotion have typically created their own stimuli sets. As such, a single group of standard materials has not been widely adopted. To cover the spectrum of materials used in emotional prosody research (i.e., for speech: words, lexical and neutral sentences; pseudo-speech: pseudo-words/sentences; non-speech: affect bursts) a total of 1038 stimuli was selected from established databases (i.e., *Berlin Database of Emotional Speech*, *Magdeburg Prosody Corpus* and *Montreal Affective Voices*) or from researchers who provided their own stimulus material for the purposes of this research project. The extracted stimuli were normalized and validated in a fairly large sample size ($N = 290$ participants) in order to provide an up-to-date set of prosodic stimuli that could be used for a number of specific measurement tasks in future research.

Manuscript 1: To address the gap concerning sex differences in literature, the first aim of this study was to examine whether the recognition of vocal emotions differs as a function of listeners' and speakers' gender (**Chapter 2**). We used a wide variety of stimuli, a fairly large number of speakers, as well as, a gender-balanced sample. This allowed us to address some of the methodological concerns raised by previous investigators (e.g., Bąk, 2016; A. E. Thompson & Voyer, 2014) regarding the impact these factors might have had on the magnitude of previously reported sex differences.

Manuscript 2: A second aim of this study was to investigate the extent emotion recognition and confidence judgements are predicted by vocal stimulus type and acoustic pa-

rameters (**Chapter 3**). Several acoustic parameters that signal various emotional states are well-documented in prosody research. However, there are still debates regarding what set of cues reliably discriminate among vocal emotions (e.g., Eyben et al., 2016). Some investigators report only on the 'basic' three paralinguistic attributes of prosody (pitch, intensity and duration) and in some cases their related acoustic parameters (e.g., Paulmann et al., 2008). With the proliferation of machine learning algorithms, the general tendency among researchers is to extract and analyze as many acoustic features as possible from the speech signal. This approach, however, comes at the cost of serious difficulties in their interpretation (Eyben, Batliner, & Schuller, 2010). Using different types of stimuli, different sets of acoustic parameters render comparisons across studies exceedingly difficult. To alleviate some of these methodological concerns, we implemented a baseline set of acoustic parameters for each- and across all stimuli types. This approach allowed us to systematically analyze their influence on emotion recognition and listeners' confidence judgements (a largely neglected aspect within the vocal emotion literature).

Study 2: Inter-individual differences in males' ability to recognize emotions expressed by voices and faces

General information: Previous research reports that males are less accurate than females at recognizing emotions (e.g., A. E. Thompson & Voyer, 2014). Different explanations have been proposed for the female advantage in the recognition of nonverbal expressions. Many of these explanations refer to the different social roles, status positions of men and women, or to the biological competence of women to read others' emotions (e.g., Babchuk et al., 1985; Davis et al., 2012; A. H. Fischer & LaFrance, 2015). However, there is a lack of direct evidence why males tend to have more difficulties and are less accurate in recognizing emotions, since this skill is of similar importance for their social interactions and personal relationships. Therefore, the aim of this study was to systematically investigate whether variations in their ability to recognize emotions are due to the modality of stimulus presentation (i.e., audio, visual, audio-visual), physiological- (i.e., hormones) or psycho-social factors (i.e., personality, empathy, mood, motives), as well as, gender of encoder and stimulus type[3]. In order to guard against false-discovery or undisclosed exploitation of the so-called researchers 'degrees of freedom' (Simmons, Nelson, & Simonsohn, 2011), the present study was pre-registered (https://osf.io/w2tgr/register/565fb3678c5e4a66b5582f67).

Manuscript 3: The first aim within our pre-registered study was to examine the extent hormones (i.e., testosterone; cortisol) and the modality of stimulus presentation influence males' ability to recognize emotions (**Chapter 4**). As already mentioned, the findings reported on the association between emotion recognition and steroid- or stress related hormones are largely inconsistent. Comparing the performance accuracy in different modalities of stimulus presentation in an explicit emotion recognition task by making use of a fairly large sample size (N = 282 males), as well as, following recommended techniques on the

[3]The related hypotheses for the psycho-social factors, encoder gender and stimulus type are not addressed in this dissertation. A preliminary analysis regarding these factors is presented in ***Appendix: Study 2***.

collection, storage and analysis of saliva samples (see Kordsmeyer, Lohöfener, & Penke, 2019; Schultheiss, Dlugash, & Mehta, 2019, for details) allowed us to address some of the methodological flaws in previous investigations.

Outlook

While the first part of this dissertation addresses theoretical and methodological considerations alongside empirical findings on emotion recognition (**Chapter 1**), the second part opens the experimental section, where the three manuscripts (**Chapter 2** to **4**) from the two studies conducted will be separately discussed. The last part (**Chapter 5**) summarizes all important results in a general discussion. In addition, limitations and implications for future research will be discussed. The dissertation will close with conclusions gained from this empirical work.

Chapter 2

Gender Differences in the Recognition of Vocal Emotions

Abstract

The conflicting findings from the few studies conducted with regard to gender differences in the recognition of vocal expressions of emotion have left the exact nature of these differences unclear. Several investigators have argued that a comprehensive understanding of gender differences in vocal emotion recognition can only be achieved by replicating these studies while accounting for influential factors such as stimulus type, gender-balanced samples, number of encoders, decoders and emotional categories. This study aimed to account for these factors by investigating whether emotion recognition from vocal expressions differs as a function of both listeners' and speakers' gender. A total of $N = 290$ participants were randomly and equally allocated to two groups. One group listened to words and pseudo-words, while the other group listened to sentences and affect bursts. Participants were asked to categorize the stimuli with respect to the expressed emotions in a fixed-choice response format. Overall, females were more accurate than males when decoding vocal emotions, however, when testing for specific emotions these differences were small in magnitude. Speakers' gender had a significant impact on how listeners' judged emotions from the voice. The group listening to words and pseudo-words had higher identification rates for emotions spoken by male than by female actors, whereas in the group listening to sentences and affect bursts the identification rates were higher when emotions were uttered by female than male actors. The mixed pattern for emotion-specific effects, however, indicates that, in the vocal channel, the reliability of emotion judgments is not systematically influenced by speakers' gender and the related stereotypes of emotional expressivity. Together, these results extend previous findings by showing effects of listeners' and speakers' gender on the recognition of vocal emotions. They stress the importance of distinguishing these factors to explain recognition ability in the processing of emotional prosody. [1]

Keywords: Gender Differences, Emotion Recognition Accuracy, Voice, Speech-embedded Emotions, Affect Bursts.

[1] Lausen, A., & Schacht, A. (2018). Gender differences in the recognition of vocal emotions. *Frontiers in Psychology*, *9*, 882. doi: 10.3389/fpsyg.2018.00882

2.1 Introduction

The ability to accurately perceive the emotional states of others is a fundamental socio-cognitive ability for the successful regulation of our interpersonal relationships (A. H. Fischer & Manstead, 2008; Levenson & Ruef, 1992) and it relies on the integration of several information cues such as facial expressions, tone of voice (prosody), words or body language (Jessen & Kotz, 2011; Van den Stock, Righart, & De Gelder, 2007). Although there is a consensus among researchers that the recognition of emotions is facilitated by the availability of additional sensory channels (De Gelder & Vroomen, 2000; Klasen, Kreifelts, Chen, Seubert, & Mathiak, 2014; Paulmann & Pell, 2010), it has also been shown that using just one channel (e.g., the voice) is more than sufficient at deciphering a person's emotional state well above chance (Apple & Hecht, 1982; Jürgens, Fischer, & Schacht, 2018; Juslin & Laukka, 2001; W. F. Thompson & Balkwill, 2006, 2009).

The voice is a highly complex tool of communication or, as already Darwin (1872, 1998) pointed out, the most indicative of an individual's emotional state. Our voice discloses information not only about our biological, psychological or social status (e.g., Azul, 2013) but also expresses emotions using different domains such as prosody, semantics or non-speech sounds (i.e., affect bursts; e.g., Kraus, 2017; Schwartz & Pell, 2012). Several studies have demonstrated that the main and most obvious function of prosody and non-speech sounds is that of facilitating interaction and communication (see for example, Belin, 2006; Belin, Bestelmeyer, Latinus, & Watson, 2011; J. Fischer & Price, 2017; Hawk et al., 2009; Paulmann et al., 2012; Pell et al., 2015, for details). One of the methodological challenges when studying prosody in human speech is how to isolate processes related to the *encoding* (expressing) and *decoding* (judging) of emotions from those of processing semantic information carried by, for example, words or sentences. To circumvent this problem, researchers used either pseudo-speech or affect bursts (e.g., simulated laughter, crying) as stimulus material. While the former captures the pure effects of emotional prosody independent of lexical-semantic cues, the latter has been argued to have an adaptive value (J. Fischer & Price, 2017) and to be an ideal tool when investigating the expression of emotional information when there is no concurrent verbal information present (Pell et al., 2015).

In the context of nonverbal communication (e.g., vocal affect, facial expressions, body language), gender has been repeatedly proposed as an important factor that might influence the accuracy of performance in emotion recognition tasks (e.g., A. Fischer & Evers, 2013; Forni-Santos & Osório, 2015; Hall, 1978, 2006; Hall, Carter, & Horgan, 2000; Sokolov, Krüger, Enck, Krägeloh-Mann, & Pavlova, 2011; A. E. Thompson & Voyer, 2014). One can distinguish two major lines of research. One line assumes that females and males differ in their emotionality, personality, abilities, attitudes or behavioral tendencies (*gender differences hypothesis*; Gray, 1992) and that women are "emotional experts", more inclined to pay attention to their own and others' feelings and intuitions (Hess et al., 2000; Shields, 2002; Timmers, Fischer, & Manstead, 2003). Several studies have shown that both genders differ in the way they express (e.g., Barrett & Bliss-Moreau, 2009; McDuff, Kodra, el Kaliouby, & LaFrance, 2017; Parkins, 2012), experience (e.g., Šolcová & Lačev, 2017),

and decode or encode emotions with females outperforming males when completing tasks designed to measure non-verbal communication ability (e.g., Adams, 2012; Ambady & Rosenthal, 1998; Wells, Gillespie, & Rotshtein, 2016; Wingenbach et al., 2018; Zuckerman, Lipets, Koivumaki, & Rosenthal, 1975). In addition, meta-analytic reviews, summarizing work on gender differences concerning the ability to recognize non- verbal expressions of emotion, also reported a female advantage for emotion recognition tasks with effect sizes ranging from small to medium (e.g., Hall, 1984; McClure, 2000). Explanations for these gender-based behavior patterns range from socio-cultural influences and psychological dispositions to evolutionary perspectives (see Briton & Hall, 1995; Brody, 1997; Davis et al., 2012; Eagly & Wood, 1999, for more detailed explanations). For instance, it has been suggested that females, due to their responsibility for child rearing, are expected to be prosocial and nurturing and, thus, more responsive and accurate in judging other people's emotions (Babchuk et al., 1985; Hall, 1984; Schirmer, 2013).

Conversely, the other line of research has emphasized the homogeneity between genders across various domains (e.g., non-verbal communication, social and personality variables, psychological well-being) based on evidence from meta-analyses. For instance, Richard, Bond Jr, and Stokes-Zoota (2003) examined gender differences across domains by using a second *order meta-analysis* (see F. L. Schmidt & Oh, 2013; Zell & Krizan, 2014, for details) to characterize the average difference between males and females. With regard to nonverbal communication the authors aggregated the data from a series of experiments conducted by Rosenthal and DePaulo (1979) and found that the correlation coefficients between genders were small, ranging from $r = 0.16$, for facial cues, $r = 0.11$, for body cues to $r = 0.06$, for vocal cues. Furthermore, Hyde (2005, 2014) observed 78% of effect sizes to be small or close to zero, leading her to conclude that in many cases females and males are rather similar on most psychological dimensions (*gender similarity hypothesis*). The results of these meta-analytic reviews are useful for estimating the overall magnitude and variability of female-male comparisons across various domains. However, this line of research might under-interpret the differences between females and males for emotion recognition by failing to consider modality specific effects (Abelson, 1985; A. E. Thompson & Voyer, 2014). A comprehensive conclusion cannot be drawn when the vast majority of evidence comes from studies that assess gender effects mainly within only one modality (e.g., Hyde, 2005) or by employing only one test (*Profile of Nonverbal Sensitivity*, Rosenthal & DePaulo, 1979) to assess performance accuracy for decoding nonverbal cues (e.g., Richard et al., 2003). Thus, until further evidence on the similarities and differences between genders within specific sensory modalities is provided, the direction of these effects remains an open question.

Contrary to the growing field of research examining gender effects in the recognition of emotions within the visual modality, where researchers are working toward improving methodology by either including facial expressions with varying intensity (Wingenbach et al., 2018), dynamically rising expressions (e.g., Recio, Schacht, & Sommer, 2014; Recio, Sommer, & Schacht, 2011), or different stimulus types such as avatars, human faces or icons (A. H. Fischer, Kret, & Broekens, 2018), the investigation of these effects within the vocal domain is still understudied. This paucity persists despite a common consensus

that the voice is an important source of social information (e.g., Latinus & Belin, 2011; Morningstar, 2017). Research comparing auditory, visual, and audio-visual modalities reported significant main effects of gender (K. R. Scherer & Scherer, 2011), with females outperforming males in all three conditions of stimulus presentation (Collignon et al., 2010). Similarly, Lambrecht et al. (2014) demonstrated a significant female advantage in emotion recognition which was however restricted to vocal emotions. A female advantage was also found in studies investigating emotion recognition purely within the vocal domain (e.g., Demenescu, Kato, & Mathiak, 2015; Paulmann & Uskul, 2014; K. R. Scherer et al., 2001; Toivanen, Väyrynen, & Seppänen, 2005). These findings were corroborated by Keshtiari and Kuhlmann (2016), who investigated how gender affects the recognition of vocal expressions of emotion. Participants listened to sentences spoken in five different emotions (angry, disgust, fear, happiness, and sadness) or in a neutral tone of voice and made a decision on the emotional category the presented utterances corresponded to. Results revealed a significant main effect of gender with an overall recognition advantage for females, confirming in this way the consistency of findings in past research. Other studies, however, reported either only a small overall advantage in favor of females in the recognition of non-verbal (auditory, visual, audio-visual) displays of emotion (Kret & De Gelder, 2012; A. E. Thompson & Voyer, 2014) or even equal performance accuracy for male and female participants in identifying emotions from both, speech-embedded (e.g., Paulmann et al., 2008; Raithel & Hielscher-Fastabend, 2004; Sauter, Panattoni, & Happé, 2013) and non-speech sounds (e.g., Hawk et al., 2009; Lima, Alves, Scott, & Castro, 2014).

To address these diverging findings, it has been suggested that instead of examining gender effects across emotions, specific emotion categories should be considered separately (De Gelder, 2016). For instance, in a behavioral study Bonebright et al. (1996) examined participants' ability to decode emotions from vocal cues. They instructed trained actors to record paragraph-long stories, each time using their voice to portray a specified emotion (i.e., anger, fear, happiness, sadness, and neutral). Subsequently, undergraduate students listened to each recorded paragraph and tried to determine which emotion the speaker was trying to portray. Females were significantly more accurate than males in decoding voices that expressed fear, happiness, and sadness. These gender differences were small but consistent. No gender differences were found for emotional expressions uttered in an angry or neutral tone of voice. Subsequent evidence showed that females outperform males for utterances spoken in a *fearful* (Demenescu et al., 2015; Zupan, Babbage, Neumann, & Willer, 2017), *happy* (Demenescu et al., 2015; Fujisawa & Shinohara, 2011; Lambrecht et al., 2014; Zupan et al., 2017), and *sad* (Fujisawa & Shinohara, 2011; Zupan et al., 2017) tone of voice. While both genders were found to perform equally well when identifying *angry* (Demenescu et al., 2015; Fujisawa & Shinohara, 2011; Lambrecht et al., 2014; Zupan et al., 2017), and *neutral* (Demenescu et al., 2015) prosody, other investigators failed to replicate these findings and found higher accuracy for females in correctly recognizing neutral vocalizations (Lambrecht et al., 2014), or no gender differences in the recognition of sad prosody Demenescu et al. (2015). That the accuracy of performance varies across discrete emotion categories (e.g., fear, sadness or happiness was argued to play a greater

role in women, whereas anger and disgust in men) might be the result of biological or environmental factors, which are likely to trigger "qualitatively" different emotional experiences for men and women (see Schirmer, 2013, for a comprehensive review).

The above-mentioned studies do not show a consistent gender pattern either regarding overall effects in the performance accuracy of decoding vocal emotions or emotion specific categories [see ***Table1*** (a1) for overall effects in decoding vocal emotions and (a2) for decoding performance accuracy by emotion categories]. There are several likely sources for these inconsistencies. One of the reasons may have been the large variety of different types of vocal stimuli (e.g., words, pseudo-words, sentences, pseudo-sentences, affect bursts). Other methodological differences that might have been responsible for these conflicting results are related either to the *number of emotions* studied [which vary from two (e.g., Collignon et al., 2010) to nine (e.g., Belin et al., 2008)], the *language under investigation* (e.g., Keshtiari & Kuhlmann, 2016; K. R. Scherer et al., 2001), the *population in question* [children (e.g., Fujisawa & Shinohara, 2011; Sauter et al., 2013), young adults (e.g., Paulmann & Uskul, 2014; K. R. Scherer et al., 2001), older adults (e.g., Lima et al., 2014), clinical populations (e.g., Zupan et al., 2017)], *unbalanced gender groups* [e.g., 71F/50M (Hawk et al., 2009)], and the *sample size* [which range from 24 (e.g., Raithel & Hielscher-Fastabend, 2004) to 428 (e.g., K. R. Scherer et al., 2001)].

The gender of the actor/actress portraying different emotions is a further variable of interest that has been proposed to influence the overall performance accuracy when identifying emotions from the voice (e.g., K. R. Scherer et al., 2001). In a validation study concerning the identification of vocal emotions, Belin et al. (2008) tested for differences in performance accuracy based on listeners' as well as speakers' gender. Participants were asked to evaluate actors' vocalizations on three emotional dimensions: valence, arousal, and intensity. Results showed higher mean identification rates (for intensity and arousal dimensions) across all emotion categories when spoken by female actors. Similar to other findings (e.g., Bonebright et al., 1996; Lambrecht et al., 2014), Belin et al. (2008) found no significant interaction between listeners' gender, speakers' gender and emotions, but a significant main effect for listeners' and speakers' gender. These findings indicate that females compared to males were not only better at decoding but also at identifying emotions in the female voice. Considering emotion-specific effects, it has been shown that vocal portrayals of anger and fear have higher mean identification rates when spoken by male actors (Bonebright et al., 1996), whereas happy (Bonebright et al., 1996), and neutral expressions (Young et al., 2017) were better identified from female voices. In contrast, other investigators observed that fear and disgust were better identified when spoken by a female (though a response bias toward disgust when an actor portrayed the emotion and, fear when an actress expressed the emotion was reported; see Collignon et al., 2010, for details). Further research that includes speakers' gender as an additional factor, reports that while gender differences might exist for identifying emotions from speakers' voice, these are not systematic and vary for specific emotions (Hawk et al., 2009; Pell, Kotz, Paulmann, & Alasseri, 2005) or occur regardless of the actors' gender (Riviello & Esposito, 2016; Schirmer & Kotz, 2003). Similar to the performance accuracy of decoding emotions,

the evidence with regard to speaker's gender as a relevant factor for identifying emotions from the voice is inconsistent [see ***Table1*** (b1) for overall identification rates by speakers' gender and (b2) for identification rates by speakers' gender and emotion category]. The discrepancies in these findings are likely to be attributable to a number of methodological differences, such as *recording conditions* (e.g., Burkhardt, Paeschke, Rolfes, Sendlmeier, & Weiss, 2005), *number of speakers* which vary from 2 (e.g., Demenescu et al., 2015) to 14 (Toivanen et al., 2005) or validity of prosodic stimuli derived from the *simulation of emotional expressions* (see Hawk et al., 2009; Jürgens, Grass, Drolet, & Fischer, 2015, for a discussion whether authentic vs. play acted emotional speech may lower ecological validity).

Table 1 | Gender differences: main findings of previous studies

Studies	Stimulus types	(a1) Overall effects of decoding vocal emotions	
		Female (F)	Male (M)
Bonebright et al., 1996	Short stories	↑	↓
Scherer et al., 2001; Paulmann & Uskul., 2014	Pseudo-sentences	↑	↓
Belin et al., 2008; Collignon et al., 2010	Affect bursts	↑	↓
Demenescu et al., 2015	Pseudo-words	↑	↓
Toivanen et al., 2005; Keshtiari & Kuhlmann, 2016; Zupan et al., 2016	Lexical & Neutral sentences	↑	↓
Hawk et al., 2009; Sauter et al., 2013; Lima et al., 2014	Affect bursts & three-digit numbers	*n.s.*	
Raithel & Hielscher-Fastabend, 2004; Paulmann et al. 2008	Lexical & Neutral sentences	*n.s.*	

		(a2) Decoding accuracy by emotion category													
		Ha		*An*		*Di*		*Fe*		*Sa*		*Su*		*Ne*	
		F	M	F	M	F	M	F	M	F	M	F	M	F	M
Bonebright et al., 1996	Short stories	↑	↓	*n.s.*				↑	↓	↑	↓			*n.s.*	
Fujisawa & Shinohara, 2011	Words	↑	↓	*n.s.*						↑	↓			*n.r.*	
Lambrecht et al., 2014	Words	↑	↓	*n.s.*		*n.s.*								↑	↓
Demenescu et al., 2015	Pseudo-words	↑	↓	*n.s.*		*n.s.*		↑	↓	*n.s.*				*n.s.*	
Zupan et al., 2016	Neutral sentences & short stories	*n.s.*		*n.s.*				↑	↓	↑	↓				

		(b1) Overall identification rates of vocal emotions by the gender of encoder	
		Female (F)	*Male (M)*
Scherer et al., 2001	Pseudo-sentences	↑	↓
Belin et al., 2008; Collignon et al., 2010	Affect bursts	↑	↓
Riviello & Esposito, 2016	Audio clips	*n.s.*	
Lambrecht et al., 2014	Words	*n.s.*	

		(b2) Identification rates by encoders' gender & emotion category													
		Ha		*An*		*Di*		*Fe*		*Sa*		*Su*		*Ne*	
		F	M	F	M	F	M	F	M	F	M	F	M	F	M
Bonebright et al., 1996	Short stories	↑	↓	↓	↑			↓	↑	*n.s.*				*n.r.*	
Pell et al., 2005	Pseudo-sentences (German)	*n.s.*		*n.s.*		↓	↑	*n.s.*		↑	↓	*n.s.*		*n.s.*	
	Pseudo-sentences (English)	↑	↓	↓	↑	↓	↑	*n.s.*		*n.s.*		↑	↓	*n.s.*	
	Pseudo-sentences (Arabic)	*n.s.*		↑	↓	*n.s.*		*n.s.*		*n.s.*		*n.s.*		*n.s.*	
Collignon et al., 2010	Affect bursts					↑	↓	↑	↓						

Ha (Happy); *An* (Angry); *Di* (Disgust); *Fe* (Fear); *Sa* (Sad); *Su* (Surprise); *Ne* (Neutral). The shades indicate the absence of emotions; *n.s.* = not significant, *n.r.* = not reported.↑/↓= better/lower performance of decoding vocal emotions by listeners' gender (**a1**) & emotion category (**a2**); better/lower performance of identifying vocal emotions by speakers' gender (**b1**) & emotion category (**b2**).

A seemingly inevitable conclusion after reviewing past work on gender differences in the recognition of vocal expressions of emotion is that conflicting findings have left the exact nature of these differences unclear. Although accuracy scores from some prior studies suggest that females are overall better than males at decoding and encoding vocal emotions, independent of the stimulus type, other studies do not confirm these findings. Likewise, the question whether women are consistently better than men at decoding and identifying emotions such as happiness, fear, sadness or neutral expressions when spoken by a female, while men have an advantage for anger and disgust, remains unresolved. The absence of consistent gender effects for the encoding and decoding of emotional vocal expressions might be a result of the selected stimuli, either speech-embedded (pseudo/words, pseudo/sentences) or non- verbal vocalizations (affect bursts). Thus, it has been suggested that a comprehensive understanding of gender differences in vocal emotion recognition can only be achieved by replicating these studies while accounting for influential factors such as stimulus type, gender-balanced samples, number of encoders, decoders, and emotional

categories (Bąk, 2016; Bonebright et al., 1996; Lambrecht et al., 2014; Pell, 2002).

To address some of these limitations, the present study aimed at investigating, across a large set of speech-embedded stimuli (i.e., words, pseudo-words, sentences, pseudo-sentences) and non-verbal vocalizations (i.e., affect bursts) whether emotion recognition of vocal expressions differs as a function of both decoders' and encoders' gender and to provide parameter estimates on the magnitude and direction of these effects. To date, no extensive research on differences between males and females in the recognition of emotional prosody has been conducted and, thus, we based our approach for investigating these effects on the patterns observed in the majority of the aforementioned studies. We first examined whether there are any differences in the performance accuracy of decoding vocal emotions based on listeners' gender (i.e., across all stimuli and for each stimulus type; across all emotions and for each emotion category). Specifically, we expected an overall female advantage when decoding vocal emotions, and that they would be more accurate than males when categorizing specific emotions such as *happiness*, *fear*, *sadness*, or *neutral* expressions. No gender differences were expected to manifest for emotions uttered in an *angry* and *disgusted* tone of voice. Secondly, we tested whether there are any differences for identifying vocal emotions based on speakers' gender (i.e., across all stimuli and for each stimulus type; across all emotions and for each emotion category). We hypothesized that vocal portrayals of emotion would have overall significantly higher hit rates when spoken by female than by male actors. Considering emotion- specific effects, we expected that *anger* and *disgust* would have higher identification rates when spoken by male actors, whereas portrayals of *happiness*, *fear*, *sadness*, and *neutral* would be better identified when spoken by female actors. Finally, we investigated potential interactions between listeners' and speakers' gender for the identification of vocal emotions across all stimuli and for each stimulus type.

2.2 Methods

The study was conducted in accordance with the ethical principles formulated in the *Declaration of Helsinki* and approved by the ethics committee of the *Georg-Elias-Mueller-Institute of Psychology*, University of Goettingen, Germany.

2.2.1 Participants

Participants were $N = 302$ volunteers (age range 18-36) from the University of Goettingen and the local community. They were recruited through flyers distributed at the University campus, the *ORSEE* database for psychological experiments (`http://www.orsee.org/web/`), postings on the social media site *Facebook* and the online platform *Schwarzes Brett* Goettingen (`https://www.uni-goettingen.de/en/644.html`). Inclusion criteria for participation in the study were: native speakers of German, aged above eighteen, normal hearing, not currently taking medication affecting the brain and no self-reported mental health problems. Twelve participants who reported hearing disorders (e.g., tinnitus), psychiatric/neurological disorders or the intake of psychotropic medication were not eligible

to participate. This left a total of 290 participants (143 female, 147 male) with a mean age of 23.83 years (SD = 3.73). To assess the performance accuracy between females and males within different types of vocal stimuli (i.e., words, pseudo-words, sentences, pseudo-sentences, affect bursts) and to reduce the length of the experiment participants were randomly allocated to two groups of equal size. This allowed us to have a higher number of stimuli in each group resulting in a higher precision of estimated gender or emotion differences within one database and respectively within one of the groups. One group classified words and pseudo-words stimuli (n = 145, M_{age} = 24.00, SD = 3.67), whereas the other group was presented with stimuli featuring sentences, pseudo-sentences and affect bursts (n = 145, M_{age} = 23.66, SD = 3.80). To assess whether there were any age differences in the two groups a Wilcoxon-Mann-Whitney test was conducted. The results indicated a significant age difference between females and males in both groups ($Group_{\text{Words}}$: z = -2.91, p = .004; $Group_{\text{Sentences}}$: z = -2.79, p = .005). Participants' demographic characteristics are presented in ***Table2***. Throughout the article these two groups will be referred to as $Group - Words$ and $Group - Sentences$. Participants were reimbursed with course credit or 8 Euros.

Table 2 | Demographic characteristics of the study population

Group	Gender		Age	Education			
		n	*M (SD)*	*HS-Dipl.*	*A-levels*	*BA*	*MA*
Words	Females	71	23.10 (3.31)	1	46	21	3
	Males	74	24.86 (3.80)	1	45	20	8
Sentences	Females	72	22.72 (3.29)		46	20	6
	Males	73	24.57 (4.06)		44	12	17

HS-Dipl. = Highschool diploma (i.e., *Realschulabschluss*); *BA* = Bachelor; *MA* = Master

Materials and Stimuli selection

The speech/non-speech embedded stimuli were extracted from well-established and validated databases or provided by researchers who developed their own stimulus materials [see ***Table 3*** for a brief description on the features of the selected databases (e.g., stimuli types, number of speakers)].

To be included in the present study the stimuli had to satisfy the following criteria: (1) be spoken in a neutral tone (i.e., baseline expression) or in one of the emotion categories of interest (i.e., happiness, surprise, anger, fear, sadness, disgust), (2) to be recorded under standardized conditions, (3) to have at least two encoders (i.e., male/female) and (4) to be produced by human expressers.

We decided to use a wide variety of stimuli representing the spectrum of materials used in emotional prosody research (i.e., for speech: words, lexical and neutral sentences; pseudo-speech: pseudo-words/sentences; for non-speech: vocalizations). For economic reasons, only a sub-set of stimuli from each database was selected. For *Anna* and *Montreal Affective Voices (MAV)* databases all speakers for the emotion category of interest were chosen. This resulted in a total number of 88 Stimuli for *Anna* [4 Emotions (anger, happiness, sadness, neutral) x 22 Speakers] and 70 Stimuli for *MAV* [7 Emotions (anger, disgust, fear, happiness, sadness, surprise, neutral) x 10 Speakers]. The stimuli from the remaining other three databases were ordered randomly and the first 10 items per database were

selected. Stimulus selection resulted in a total number of 280 stimuli from the *Paulmann Prosodic Stimuli* set [10 *Pseudo-sentences* x 7 Emotions (anger, disgust, fear, happiness, sadness, surprise and neutral) x 2 Speakers; 10 *Lexical Sentences* x 7 Emotions (anger, disgust, fear, happiness, sadness, surprise and neutral) x 2 Speakers], 120 stimuli from the *Berlin Database of Emotional Speech* [10 *Semantic Neutral Sentences* x 6 Emotions (anger, disgust, fear, happiness, sadness and neutral) x 2 Speakers] and 480 Stimuli from the *Magdeburg Prosody Corpus* [10 *Pseudo-words* x 6 Emotions (anger, disgust, fear, happiness, sadness and neutral) x 2 Speakers; 10 *Semantic positive nouns*/ 10 *Semantic negative nouns*/ 10 *Semantic neutral nouns* x 6 Emotions (anger, disgust, fear, happiness, sadness and neutral) x 2 Speakers]. The nouns extracted from the *Magdeburg Prosody Corpus* were additionally controlled for valence, arousal and word frequency according to the *Berlin Affective Word List Reloaded* (Võ et al., 2009).

Table 3 | Features of the selected emotion speech databases[a]

Database	Speakers	Emotions	Nature of material	Total stimuli
Anna (Hammerschmidt & Jürgens, 2007)	22 drama students (10 male/12 female)	Anger, affection, contempt, despair, fear, happiness, sensual satisfaction, triumph, neutral	Word	$N_{Stimuli}$ = 198
Berlin Database of Emotional Speech (EMO_DB) (Burkhardt et al., 2005)	10 untrained actors (5 male/5 female)	Anger, boredom, disgust, fear, happiness, sadness, neutral	Semantic neutral sentences	$N_{Stimuli}$ = 816
Magdeburg Prosody Corpus (WASEP) (Wendt & Scheich, 2002)	2 actors (1 male/1female)	Anger, disgust, fear, happiness, sadness, neutral	Pseudo-words	$N_{Stimuli}$ = 222
			Nouns[b]	$N_{Stimuli}$ = 3318
Montreal Affective Voices (MAV) (Belin et al., 2008)	10 actors (5 male/5 female)	Anger, disgust, fear, happiness, pain, pleasure, sadness, surprise, neutral	Affect bursts	$N_{Stimuli}$ = 90
Paulmann Prosodic Stimuli (Paulmann & Kotz, 2008; Paulmann et al., 2008)	2 actors (1 male/1female)	Anger, disgust, fear, happiness, sadness, surprise, neutral	Pseudo- sentences	$N_{Stimuli}$ = 210
			Lexical sentences[c]	$N_{Stimuli}$ = 210

[a]The word databases it is used as a generic term as some of the selected stimuli are from researchers that developed their own stimulus materials with no aim of establishing a database (i.e., Anna & Paulmann prosodic stimuli). [b]The nouns from *WASEP* are classified according to their *positive*, *negative* and *neutral* semantic content. [c]*Paulmann lexical sentences* consists of semantically and prosodically matching stimuli. Compared to all other types of stimuli, which were cros-sover designed (i.e., stimulus is spoken in all emotional categories) both, the *pseudo-* and *lexical sentences* from Paulmann et al. (2008) database were hierarchically designed (i.e., stimulus is spoken only in one emotional category). The validation procedures of the stimuli are presented in the above-cited papers.

2.2.2 Acoustic analysis

The extraction of amplitude (dB), duration and peak amplitude of all 1038 original stimuli was conducted using the phonetic-software Praat (Boersma, 2001). As the stimuli used for this study came from different databases with different recording conditions, we controlled for acoustic parameters, including the minimum, maximum, mean, variance, and standard deviation of the amplitude. The results of our analyses indicated that the variation coefficient (C_V) for amplitude between the stimuli was high (s^2 = 71.92, M = 63.06, C_V = 13.45%). Therefore, the stimuli were normalized with regards to loudness by applying the *Group Waveform Normalization* algorithm of *Adobe Audition CC* (Version 8.1, Adobe Systems, 2015, San Jose, CA) that uniformly matches the loudness based on the *root-mean-square* (RMS) levels. To control whether normalization worked, the stimuli were re-uploaded in *Praat*, which indicated that the variation coefficient between the stimuli was reduced by roughly 40% (s^2 = 24.97, M = 61.07, C_V = 8.18%) by this procedure.

Physical volume of stimulus presentation across the four PCs' used in the experiment was controlled by measuring sound volume of the practice trials with a professional sound level meter, *Nor140* (Norsonic, 2010, Lierskogen, Norway). No significant difference in volume intensity was observed [$F_{(3,27)}$ = 0.53, p = .668].

2.2.3 Procedure

Participants were tested in groups of up to four members. At arrival, each participant was seated in front of a *Dell OptiPlex^{TM} 780* Desktop-PC. All participants were provided with

individual headphone devices (*Bayerdynamic DT 770 PRO*). After signing a consent form and completing a short demographic questionnaire concerning age, gender[2] and education level, participants were informed that they would be involved in a study evaluating emotional aspects of vocal stimulus materials. Afterwards, they were told to put on headphones and carefully read the instructions presented on the computer screen. Before the main experiment, participants were familiarized with the experimental setting in a short training session comprised of 10 stimuli, which were not presented in the main experiment. They were instructed to carefully listen to the presented stimuli as they would be played only once and that the number of emotions presented might vary form the number of categories given as possible choices (see Design & Randomization for an argument related to this approach). Each trial began with a white fixation-cross presented on a grey screen, which was shown until participants' response had been recorded. The presentation of the stimuli was initiated by pressing the *Enter*-key. After stimulus presentation, participants had to decide as accurately as possible, in a fixed-choice response format, which of the 7 emotional categories (i.e., anger, disgust, fear, happiness, sadness, surprise, neutral) the emotional prosody of the presented stimulus corresponded to. Following their emotion judgment, they were asked the correctness of their answer on a 7-point Likert scale, where '1' corresponded to *not at all confident* and '7' corresponded to *extremely confident*. The responses were made using the marked computer keyboard (*Z* to *M* for the emotion judgments, which were labeled corresponding to the emotion categories, and *1* to *7* for confidence). There was no time limit for emotion judgments or confidence ratings. At the end of each block a visual message in the center of the screen instructed participants to take a break if they wished to or to press the *Spacebar* to proceed with the next block. The 568 stimuli for *Group Words* had a mean duration of 1.03 ± 0.36 s, whereas in *Group Sentences* the mean duration of the 470 stimuli was 2.66 ± 1.01 s. Testing took approximately 60 minutes for both groups.

2.2.4 Design & Randomization

We fitted a balanced design to allow for a separate analysis of effects across the recognizability of emotional expressions, skill in judging emotional expressions, and the interaction between encoding and decoding (the assumptions of such an approach were justified by Elfenbein & Ambady, 2002a, 2002b). Following the argumentation of Wagner (1993), participants were provided with the same number of judgment categories, independent of the given emotion categories within the included databases. This approach guarantees that, the response probabilities are not influenced by the different number of emotional categories (i.e., the probability of correct/false recognition of emotions by random choice is equal).

The set of stimuli for the *Group Words* was split into three blocks (*Anna*, *Pseudo-words* and *Nouns*) while the set of stimuli for the *Group Sentences* was split into four blocks

[2]We decided to use the term "gender" instead of "sex" because this concept has a wider connotation and is not purely assigned by genetics. Participants had the option to provide a textual answer with regard to their gender ("Geschlecht"), yet all of them identified either as male or female (none of them wrote "sexless," "I am born male but feel female" etc.).

(*Pseudo-sentences*, *Lexical Sentences*, *Neutral sentences* and *Affect bursts*). Each block as well as the stimuli within each block were randomized using the software *Presentation* (Version 14.1, Neurobehavioral Systems Inc., Albany, CA).

2.2.5 Sample size calculations

A target sample size of 134 participants per group (67F/67M) was determined using *Wilcoxon Mann-Whitney* two-tailed test ($d = .50$; $\alpha = .05$; $1 - \beta = .80$). Assuming 67 participants in each gender group and the minimum number of observations per participant (i.e., 70) we further investigated, via a two-sample binomial test, whether the determined sample size possessed enough statistical power to assess the size of females'/males' differences in detecting vocal emotions. This argument indicated that at 80% recognition probability the sample size was powered enough to detect small differences as 2.3%. To take account of possible attrition the sample size was increased by at least 10%.

2.2.6 Statistical Analysis

The data was analyzed by a *generalized linear model* (*quasi-binomial logistic regression*) for the binary response variable emotion recognition. As individual effects (e.g., fatigue, boredom) might impact cognitive performance, we treated *participants* as a confounder in our model. In addition, we controlled for *confidence*, as a confounder, shown to impact on performance accuracy in emotion recognition tasks (e.g., Beaupré & Hess, 2006; Rigoulot, Wassiliwizky, & Pell, 2013). Our analysis on the baseline characteristics of the study population indicated a significant age effect between males and females and, therefore, we additionally included this factor as a confounder in our model. Listeners' gender, speakers' gender, emotions and stimulus type were included as predictor variables. Age was included as a quantitative variable. Listeners' gender and speakers' gender were included as binary variables and confidence, participant, emotion and stimulus type were included as nominal variables. The dispersion parameter of the quasi-binomial model and the nominal variable participants accounted for dependencies caused by repeated measurements within the participants. First order interactions were fitted between *listeners' gender* and *speakers' gender*, *age* and *stimuli types*, *confidence* and *stimuli types*, *speakers' gender* and *stimuli types*, *listeners' gender* and *stimuli types*, *emotions* and *stimuli types*, *age* and *speakers' gender*, *age* and *emotions*, *listeners' gender* and *emotions*, *speakers' gender* and *emotions*, *confidence* and *participant*. A second order interaction was fitted between *listeners' gender*, *speakers' gender* and *emotions*. Chi-square tests of the deviance analysis (generalized mixed model analysis) were used to analyze additive and interaction effects.

Means, standard deviations, z-scores, p-values and effect sizes were calculated to describe the differences between genders in performance accuracy. This descriptive analysis was conducted using the unadjusted group means, which allows the application of non-parametric robust methods, direct illustration and interpretation of the effect sizes and patterns. As emotion recognition is binomial distributed and does not allow the assumption of a normal distribution we used *Wilcoxon-Mann-Whitney test* for independent samples to analyze the effects of listeners' gender and Wilcoxon-rank-sum test for dependent

samples to analyze the effects of speakers' gender. Corrections for multiple testing were implemented using Bonferroni's method for multiple comparisons.

The data was analyzed using the R language and environment for statistical computing and graphics version 3.3.1 (R Core Team, 2017) and the integrated environment R-Studio version 1.0.316. The quasi-binomial logistic regression was fitted using the R function glm. *Wilcoxon-Mann-Whitney* and *Wilcoxon-rank-sum* were performed with the R package *coin* introduced by Hothorn, Hornik, Van De Wiel, Zeileis, et al. (2008).

2.3 Results

2.3.1 Emotion Effects on Performance Accuracy

The quasi-binomial logistic models revealed significant interactions between emotions and stimuli types in both groups (*Group Words*: $\chi^2_{(18)} = 1097.80$, $p < .001$; *Group Sentences*: $\chi^2_{(17)} = 1990.40$, $p < .001$). Main effects of emotion were observed across all stimuli (*Group Words*: $\chi^2_{(5)} = 4853.80$, $p < .001$; *Group Sentences*: $\chi^2_{(6)} = 6956.00$, $p < .001$) and for each stimulus type [*Anna* ($\chi^2_{(3)} = 2463.87$, $p < .001$), *pseudo-words* ($\chi^2_{(5)} = 1060.19$, $p < .001$), *semantic positive nouns* ($\chi^2_{(5)} = 616.96$, $p < .001$), *semantic negative nouns* ($\chi^2_{(5)} = 735.54$, $p < .001$) *semantic neutral nouns* ($\chi^2_{(5)} = 1603.56$, $p < .001$), *pseudo-sentences* ($\chi^2_{(6)} = 2784.06$, $p < .001$), *lexical sentences* ($\chi^2_{(6)} = 3745.60$, $p < .001$), *neutral sentences* ($\chi^2_{(5)} = 1332.93$, $p < .001$) and *affect bursts* ($\chi^2_{(6)} = 1113.20$, $p < .001$)]. The full models across all stimuli and for each stimulus type are presented in Supplementary material (see *Tables S1a,b; S2a-i*).

2.3.2 Decoding Performance Accuracy by Listeners' Gender

Significant first order interactions between listeners' gender and stimuli types were observed for both groups [*Group Words* ($\chi^2_{(4)} = 16.40$, $p = .038$); *Group Sentences* ($\chi^2_{(3)} = 22.80$, $p < .001$)]. A significant main effect of gender was found across the stimuli types for *Group Words* ($\chi^2_{(1)} = 51.70$, $p < .001$) but not for *Group Sentences* ($\chi^2_{(1)} = 5.20$, $p = .332$). Main effects of gender were revealed for the following stimulus sub-sets: *pseudo-words* ($\chi^2_{(1)} = 29.18$, $p < .001$), *semantic positive nouns* ($\chi^2_{(1)} = 20.14$, $p < .001$), *semantic negative nouns* ($\chi^2_{(1)} = 8.38$, $p = .046$) and *semantic neutral nouns* ($\chi^2_{(1)} = 9.13$, $p = .029$). No main effects of gender were found for *Anna* ($\chi^2_{(1)} = 0.03$, $p = 1.00$), *pseudo-sentences* ($\chi^2_{(1)} = 7.44$, $p = .068$), *lexical sentences* ($\chi^2_{(1)} = 5.50$, $p = .211$), *neutral sentences* ($\chi^2_{(1)} = 5.35$, $p < .243$) and *affect bursts* ($\chi^2_{(1)} = 2.83$, $p = .698$).

The quasi-binomial logistic models showed significant first order interactions between listeners' gender and emotions for both groups (*Group Words*: ($\chi^2_{(5)} = 26.60$, $p < .001$; *Group Sentences*: ($\chi^2_{(6)} = 19.60$, $p = .029$). When testing the performance accuracy by stimulus type there were no significant interactions between listeners' gender and emotions [*Anna* ($\chi^2_{(3)} = 7.61$, $p = .644$), *pseudo-words* ($\chi^2_{(5)} = 15.18$, $p = 0.117$), *semantic positive nouns* ($\chi^2_{(5)} = 6.73$, $p = 1.00$), *semantic negative nouns* ($\chi^2_{(5)}= 14.96$, $p = .124$), *semantic neutral nouns* ($\chi^2_{(5)} = 3.12$, $p = 1.00$), *pseudo-sentences* ($\chi^2_{(6)} = 17.36$, $p = .075$), *lexical sentences* ($\chi^2_{(6)} = 9.60$, $p = 1.00$), *neutral sentences* ($\chi^2_{(5)} = 4.05$, $p = 1.00$) and *affect bursts*

($\chi^2_{(6)} = 5.90$, $p = .848$)]. ***Figure 1*** **(A,B)** illustrates the performance accuracy by listeners' gender and emotion categories, separated for both, *Group-Words* and *Group-Sentences*.

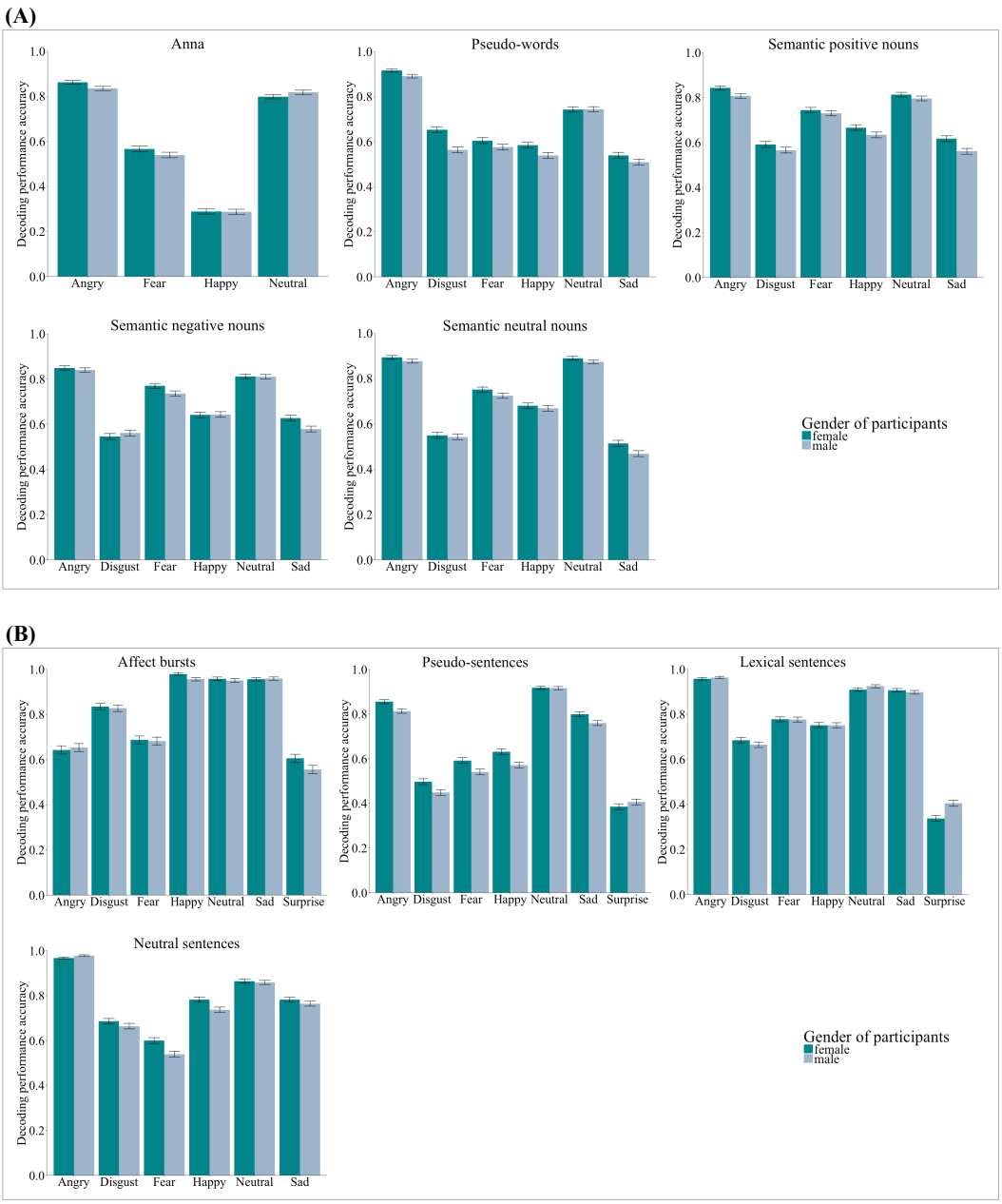


Figure 1 | Bar charts showing the performance accuracy by listeners' gender [(**A**) *Group Words* (*n* = 145, 71 females); (**B**) *Group Sentences* (*n* = 145, 72 females)]. Error bars represent the standard error. As it can be observed, for the majority of emotion categories by databases, females had higher decoding performance accuracy than males.

To describe the size of the difference between females and males when decoding emotions, a *Wilcoxon Mann-Whitney* test was implemented. Results showed that overall females ($M = 0.67$, $SD = 0.06$) were significantly better than males ($M = 0.64$, $SD = 0.06$) at decoding emotions from pseudo-sentences, $z = 2.87$, $p = .033$, $d = .49$. No significant differences between females and males by emotion category were observed, with effect sizes close to zero ($d \leq 0.10$) or in the small ($0.11 < d < 0.35$) range. Overall, the results indicated the existence of a small decoding effect favoring females across all emotions (0.15

$< d < 0.34$) and stimulus types ($d = 0.31$). The parameter estimates by listeners' gender for each emotion category, across all emotions and stimulus types are presented in ***Table 4***.

Table 4 | Means, standard deviations, z-scores, p-values and effect sizes of performance accuracy by listeners' gender

	Stimulus type	Emotion category	Female		Male				
			n	*M (SD)*	*n*	*M (SD)*	*z*	*p*	*d*
Group Words	Anna	*Angry*	71	0.86 (0.09)	74	0.83 (0.12)	1.26	1.00	0.26
		Fear	71	0.57 (0.16)	74	0.54 (0.15)	1.03	1.00	0.17
		Happy	71	0.29 (0.11)	74	0.29 (0.11)	-0.23	1.00	0.01
		Neutral	71	0.80 (0.16)	74	0.82 (0.14)	-0.43	1.00	0.12
		Overall	71	0.63 (0.06)	74	0.62 (0.06)	0.82	1.00	0.15
	Pseudo-words	*Angry*	71	0.92 (0.09)	74	0.89 (0.12)	1.48	0.967	0.24
		Disgust	71	0.65 (0.22)	74	0.56 (0.22)	2.65	0.057	0.41
		Fear	71	0.60 (0.19)	74	0.58 (0.19)	0.82	1.00	0.15
		Happy	71	0.58 (0.17)	74	0.54 (0.21)	1.27	1.00	0.23
		Neutral	71	0.74 (0.16)	74	0.74 (0.20)	-0.50	1.00	0.00
		Sad	71	0.54 (0.29)	74	0.51 (0.28)	0.65	1.00	0.11
		Overall	71	0.67 (0.11)	74	0.64 (0.12)	1.74	0.572	0.32
	Semantic positive nouns	*Angry*	71	0.84 (0.11)	74	0.81 (0.13)	1.69	.637	0.30
		Disgust	71	0.59 (0.22)	74	0.57 (0.23)	0.67	1.00	0.11
		Fear	71	0.74 (0.16)	74	0.73 (0.17)	0.65	1.00	0.08
		Happy	71	0.67 (0.14)	74	0.64 (0.15)	1.31	1.00	0.22
		Neutral	71	0.81 (0.15)	74	0.80 (0.15)	0.68	1.00	0.11
		Sad	71	0.62 (0.26)	74	0.56 (0.27)	1.28	1.00	0.21
		Overall	71	0.71 (0.08)	74	0.68 (0.09)	1.96	0.352	0.34
	Semantic negative nouns	*Angry*	71	0.85 (0.12)	74	0.84 (0.11)	0.90	1.00	0.08
		Disgust	71	0.55 (0.21)	74	0.56 (0.22)	-0.35	1.00	0.07
		Fear	71	0.77 (0.13)	74	0.74 (0.15)	1.42	1.00	0.23
		Happy	71	0.64 (0.15)	74	0.64 (0.19)	-0.60	1.00	0.01
		Neutral	71	0.81 (0.17)	74	0.81 (0.16)	0.18	1.00	0.01
		Sad	71	0.63 (0.26)	74	0.59 (0.26)	1.26	1.00	0.19
		Overall	71	0.71 (0.07)	74	0.69 (0.09)	0.62	1.00	0.15
	Semantic neutral nouns	*Angry*	71	0.89 (0.10)	74	0.88 (0.10)	1.08	1.00	0.16
		Disgust	71	0.55 (0.21)	74	0.54 (0.21)	0.24	1.00	0.03
		Fear	71	0.75 (0.17)	74	0.72 (0.18)	1.09	1.00	0.15
		Happy	71	0.68 (0.15)	74	0.67 (0.20)	-0.22	1.00	0.06
		Neutral	71	0.89 (0.13)	74	0.87 (0.14)	0.43	1.00	0.12
		Sad	71	0.51 (0.28)	74	0.47 (0.27)	0.99	1.00	0.16
		Overall	71	0.71 (0.08)	74	0.69 (0.10)	1.17	1.00	0.23
	Overall		71	0.69 (0.07)	74	0.67 (0.08)	1.39	0.163	0.31
Group Sentences	Affect bursts	*Angry*	72	0.64 (0.13)	73	0.65 (0.16)	-0.81	1.00	0.07
		Disgust	72	0.83 (0.10)	73	0.83 (0.12)	0.23	1.00	0.07
		Fear	72	0.69 (0.18)	73	0.68 (0.20)	0.07	1.00	0.03
		Happy	72	0.98 (0.05)	73	0.96 (0.11)	0.80	1.00	0.27
		Neutral	72	0.96 (0.06)	73	0.95 (0.08)	0.20	1.00	0.11
		Sad	72	0.96 (0.09)	73	0.96 (0.08)	0.24	1.00	0.04
		Surprise	72	0.61 (0.20)	73	0.56 (0.23)	1.26	1.00	0.23
		Overall	72	0.81 (0.05)	73	0.80 (0.06)	0.86	1.00	0.22
	Pseudo-sentences	*Angry*	72	0.86 (0.12)	73	0.81 (0.13)	1.92	0.434	0.32
		Disgust	72	0.50 (0.17)	73	0.45 (0.17)	1.97	0.393	0.28
		Fear	72	0.59 (0.17)	73	0.54 (0.18)	1.83	0.533	0.30
		Happy	72	0.63 (0.16)	73	0.57 (0.16)	1.99	0.368	0.37
		Neutral	72	0.92 (0.09)	73	0.92 (0.09)	-0.26	1.00	0.02
		Sad	72	0.80 (0.13)	73	0.76 (0.15)	1.74	0.649	0.28
		Surprise	72	0.39 (0.17)	73	0.41 (0.16)	-0.49	1.00	0.12
		Overall	72	0.67 (0.06)	73	0.64 (0.06)	2.87	0.033	0.49
	Lexical sentences	*Angry*	72	0.96 (0.05)	73	0.96 (0.06)	-1.19	1.00	0.10
		Disgust	72	0.69 (0.20)	73	0.66 (0.16)	0.96	1.00	0.11
		Fear	72	0.78 (0.15)	73	0.78 (0.17)	-0.29	1.00	0.01
		Happy	72	0.75 (0.16)	73	0.75 (0.17)	-0.20	1.00	0.00
		Neutral	72	0.91 (0.09)	73	0.92 (0.08)	-0.73	1.00	0.17
		Sad	72	0.91 (0.08)	73	0.90 (0.09)	0.53	1.00	0.10
		Surprise	72	0.34 (0.18)	73	0.40 (0.20)	-2.08	0.298	0.35
		Overall	72	0.76 (0.07)	73	0.77 (0.06)	-0.69	1.00	0.13
	Neutral sentences	*Angry*	72	0.97 (0.05)	73	0.98 (0.03)	-1.29	1.00	0.26
		Disgust	72	0.69 (0.14)	73	0.66 (0.17)	0.48	1.00	0.14
		Fear	72	0.60 (0.21)	73	0.54 (0.18)	1.81	0.491	0.30
		Happy	72	0.78 (0.13)	73	0.74 (0.16)	1.45	1.00	0.31
		Neutral	72	0.86 (0.13)	73	0.86 (0.13)	0.35	1.00	0.04
		Sad	72	0.78 (0.19)	73	0.76 (0.18)	0.75	1.00	0.09
		Overall	72	0.78 (0.08)	73	0.76 (0.08)	1.86	0.442	0.30
	Overall		72	0.75 (0.05)	73	0.73 (0.05)	1.60	0.110	0.31

Note: The group comparisons between males and females were made using the *Wilcoxon-Mann-Whitney* test. The decoding performance accuracy was higher for females as indicated by *positive z-scores* and higher for males as indicated by *negative z-scores*. The tests were conducted for each emotion separately, across all emotions and stimulus types. All *p-values* were Bonferroni corrected.

2.3.3 Performance Accuracy of Identifying Vocal Emotions by Speakers' Gender

The logistic regression models showed significant first order interactions between speaker gender and stimuli types [*Group Words* ($\chi^2_{(4)} = 142.80$, $p < .001$), *Group Sentences* ($\chi^2_{(3)} = 18.50$, $p = .003$)]. Main effects of speaker gender were observed across all stimuli types [*Group Words* ($\chi^2_{(1)} = 42.30$, $p < .001$), *Group Sentences* ($\chi^2_{(1)} = 589.40$, $p < .001$)] and following stimulus sub-sets: *Anna* ($\chi^2_{(1)} = 75.13$, $p < .001$), *pseudo-words* ($\chi^2_{(1)} = 22.26$, $p < .001$), *semantic negative nouns* ($\chi^2_{(1)} = 71.74$, $p < .001$), *pseudo-sentences* ($\chi^2_{(1)} = 173.65$, $p < .001$), *lexical sentences* ($\chi^2_{(1)} = 154.70$, $p < .001$) and *affect bursts* ($\chi^2_{(1)} = 40.24$, $p < .001$). No main effects of speaker gender were found for *semantic positive nouns* ($\chi^2_{(1)} = 0.43$, $p = 1.00$), *semantic neutral nouns* ($\chi^2_{(1)} = 3.05$, $p = .997$) and *neutral sentences* ($\chi^2_{(1)} = 0.93$, p = 1.00).

We observed significant first order interactions between speakers' gender and emotions across all stimuli types (*Group Words*: ($\chi^2_{(5)} = 842.30$, $p < .001$; *Group Sentences*: ($\chi^2_{(6)} = 726.70$, $p < .001$) and for each stimulus sub-set [*Anna* ($\chi^2_{(3)} = 211.41$, $p < .001$), *pseudo-words* ($\chi^2_{(5)} = 202.22$, $p < .001$), *semantic positive nouns* ($\chi^2_{(5)} = 462.14$, $p < .001$), *semantic negative nouns* ($\chi^2_{(5)} = 280.14$, $p < .001$), *semantic neutral nouns* ($\chi^2_{(5)} = 465.36$, $p < .001$), *affect bursts* ($\chi^2_{(6)} = 243.28$, $p < .001$), *pseudo-sentences* ($\chi^2_{(6)} = 1276.99$, $p < .001$), *lexical sentences* ($\chi^2_{(6)} = 194.50$, $p < .001$) and *neutral sentences* ($\chi^2_{(5)} = 449.41$, $p < .001$)]. ***Figure 2*****(A,B)** displays listeners' performance accuracy when identifying emotions from females' and males' voice.

To analyze whether specific emotions will have higher identification rates when spoken by a female than by a male encoder or vice-versa, a *Wilcoxon-rank-sum* test was fitted. Results showed that, except pseudo-sentences, in all other types of stimuli *disgust* was significantly better identified when uttered by a female than by a male (*p's* $< .001$, $0.42 < d < 2.39$). In *Group Words*, except for the name Anna, *angry* had higher identification rates in males' than females' voice (*p's* $< .001$, $0.89 < d < 1.29$), whereas in *Group Sentences* this emotion was better identified when spoken by a female than by a male (*p's* $< .001$, $0.32 < d < 1.21$). For the other emotion categories, the pattern of results was not as clear-cut: in some types of stimuli utterances were significantly better identified when spoken by female than male actors and vice-versa. Across all emotions, *Anna* ($p < .001$, $d = .84$) and *semantic negative nouns* ($p < .001$, $d = .84$) were better identified in the male voice, whereas *pseudo-words* were better identified in the female voice. No significant differences in performance accuracy when male or female actors expressed the emotions were observed for *semantic positive nouns* ($p = 1.00$, $d = .15$) and *semantic neutral nouns* ($p = .578$, $d = .18$). In *Group Sentences*, however, female utterances were significantly better identified than those spoken by male actors (*p's* $< .001$, $0.80 < d < 1.62$). Across all stimuli types, in *Group Words* vocal expressions had higher identification rates for male actors' expressions of emotion ($p < .001$, $d = 0.40$), whereas in *Group Sentences* these were better identified in the female voice ($p < .001$, $d = 2.29$). The performance accuracy by speakers' gender for each emotion category, across all emotions and stimulus types is presented in ***Table 5***.

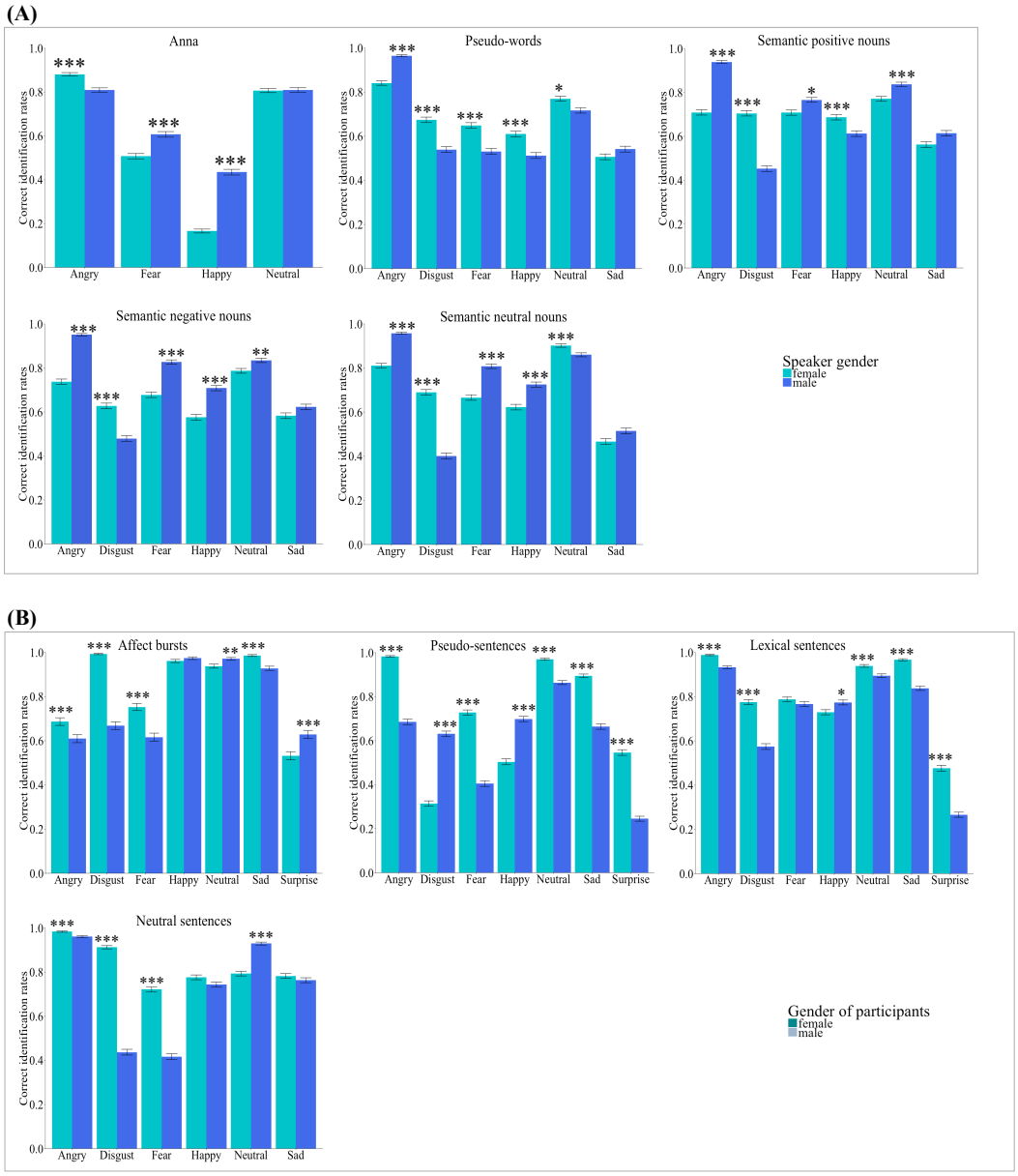


Figure 2 | Bar charts showing the performance accuracy of identifying emotions by speakers' gender. Error bars represent the standard error. Asterisks mark the significance level: *$p < .05$, **$p < .01$, ***$p < .001$. For the majority of emotion categories by databases in *Group Words* (**A**), the correct identification rates were higher for emotions uttered in a male than a female voice, while in *Group Sentences* (**B**) the opposite pattern was observed.

2.3.4 Interplay of Decoder and Encoder Gender and Emotion

No interactions between listeners and speakers gender was found across stimuli types [*Group Words* ($\chi^2_{(1)} = 0.30$, $p = 1.00$), *Group Sentences* ($\chi^2_{(1)} = 0.10$, $p = 1.00$)] or for any of the stimuli sub-sets [*Anna* ($\chi^2_{(1)} = 1.41$, $p = 1.00$), *pseudo-words* ($\chi^2_{(1)} = 0.85$, $p = 1.00$), *semantic positive nouns* ($\chi^2_{(1)} = 0.06$, $p = 1.00$), *semantic negative nouns* ($\chi^2_{(1)} = 0.25$, $p = 1.00$), *semantic neutral nouns* ($\chi^2_{(1)} = 2.80$, $p = 1.00$), *affect bursts* ($\chi^2_{(1)} = 0.21$, $p = 1.00$), *pseudo-sentences* ($\chi^2_{(1)} = 0.31$, $p = 1.00$), *lexical sentences* ($\chi^2_{(1)} = 0.20$, $p = 1.00$) and *neutral sentences* ($\chi^2_{(1)} = 1.67$, $p = 1.00$)].

The quasi-binomial logistic regression model revealed a significant second order interaction between speaker gender (encoder), listener gender (decoder) and emotion for semantic

Table 5 | Means, standard deviations, z-scores, p-values and effect sizes of identification rates by speakers' gender

	Stimulus type	Emotion category	Female		Male					
			n	*M*	*n*	*M*	*SD(Δ)*	*z*	*p*	*d*
Group Words	Anna	*Fear*	71	0.51	74	0.61	0.18	-6.01	< 0.001	0.57
		Happy	71	0.17	74	0.43	0.20	-9.72	< 0.001	1.32
		Angry	71	0.88	74	0.80	0.15	-5.41	< 0.001	0.48
		Neutral	71	0.81	74	0.81	0.17	-0.45	1.00	0.01
		Overall	71	0.59	74	0.67	0.09	-7.97	< 0.001	0.84
	Pseudo-words	*Fear*	71	0.65	74	0.53	0.25	5.21	< 0.001	0.47
		Happy	71	0.61	74	0.51	0.19	5.63	< 0.001	0.51
		Sad	71	0.51	74	0.54	0.28	-1.54	0.873	0.12
		Angry	71	0.84	74	0.96	0.13	-9.23	< 0.001	0.99
		Disgust	71	0.67	74	0.54	0.32	4.73	< 0.001	0.42
		Neutral	71	0.78	74	0.72	0.21	3.10	0.013	0.26
		Overall	71	0.68	74	0.63	0.09	4.95	< 0.001	0.44
	Semantic positive nouns	*Fear*	71	0.71	74	0.77	0.21	-3.10	0.013	0.28
		Happy	71	0.69	74	0.61	0.23	3.89	< 0.001	0.32
		Sad	71	0.56	74	0.61	0.32	-2.25	0.170	0.16
		Angry	71	0.71	74	0.94	0.18	-10.08	< .001	1.29
		Disgust	71	0.70	74	0.45	0.27	-8.65	< .001	0.93
		Neutral	71	0.77	74	0.84	0.21	-4.60	< .001	0.32
		Overall	71	0.69	74	0.70	0.09	-1.51	1.00	0.15
	Semantic negative nouns	*Fear*	71	0.68	74	0.83	0.22	-6.92	< 0.001	0.68
		Happy	71	0.58	74	0.71	0.21	-6.56	< 0.001	0.63
		Sad	71	0.58	74	0.62	0.30	-1.95	0.362	0.13
		Angry	71	0.74	74	0.95	0.19	9.74	< 0.001	1.41
		Disgust	71	0.63	74	0.48	0.29	5.50	< 0.001	0.51
		Neutral	71	0.79	74	0.83	0.18	-3.75	0.001	0.26
		Overall	71	0.66	74	0.74	0.09	-7.72	< 0.001	0.78
	Semantic neutral nouns	*Fear*	71	0.67	74	0.81	0.26	-5.80	< 0.001	0.55
		Happy	71	0.62	74	0.72	0.24	-4.82	< 0.001	0.43
		Sad	71	0.47	74	0.51	0.30	-2.14	0.227	0.16
		Angry	71	0.81	74	0.96	0.16	-8.82	< .001	0.89
		Disgust	71	0.69	74	0.40	0.29	8.53	< .001	1.00
		Neutral	71	0.90	74	0.86	0.15	3.82	< .001	0.27
		Overall	71	0.69	74	0.71	0.10	-1.74	0.578	0.18
	Overall		71	0.67	74	0.69	0.07	-4.51	< 0.001	0.40
Group Sentences	Affect bursts	*Fear*	72	0.75	73	0.62	0.28	5.25	< 0.001	0.48
		Happy	72	0.97	73	0.96	0.19	-1.09	1.00	0.12
		Sad	72	0.99	73	0.93	0.12	5.41	< 0.001	0.48
		Angry	72	0.69	73	0.61	0.24	4.52	< 0.001	0.32
		Disgust	72	0.99	73	0.68	0.22	10.28	< 0.001	1.51
		Neutral	72	0.97	73	0.94	0.13	3.24	0.009	0.26
		Surprise	72	0.53	73	0.63	0.26	-3.96	< 0.001	0.37
		Overall	72	0.84	73	0.77	0.09	8.10	< 0.001	0.80
	Pseudo-sentences	*Fear*	72	0.73	73	0.40	0.27	9.47	< 0.001	1.18
		Happy	72	0.50	73	0.70	0.23	-8.10	< 0.001	0.86
		Sad	72	0.90	73	0.66	0.25	8.50	< 0.001	0.93
		Angry	72	0.98	73	0.69	0.25	10.21	< 0.001	1.21
		Disgust	72	0.31	73	0.63	0.22	-9.76	< 0.001	1.42
		Neutral	72	0.97	73	0.86	0.12	8.60	< 0.001	0.89
		Surprise	72	0.55	73	0.25	0.25	9.43	< 0.001	1.19
		Overall	72	0.71	73	0.60	0.09	9.49	< 0.001	1.22
	Lexical sentences	*Fear*	72	0.79	73	0.77	0.22	0.72	1.00	0.10
		Happy	72	0.73	73	0.77	0.18	-2.91	0.029	0.25
		Sad	72	0.97	73	0.84	0.17	7.97	< 0.001	0.77
		Angry	72	0.99	73	0.93	0.09	7.18	< 0.001	0.64
		Disgust	72	0.78	73	0.57	0.25	7.90	< 0.001	0.82
		Neutral	72	0.94	73	0.90	0.13	4.46	< 0.001	0.33
		Surprise	72	0.48	73	0.27	0.19	9.10	< 0.001	1.10
		Overall	72	0.80	73	0.72	0.07	9.78	< 0.001	1.20
	Neutral sentences	*Fear*	72	0.72	73	0.42	0.20	10.09	< 0.001	1.48
		Happy	72	0.78	73	0.74	0.17	2.36	0.128	0.19
		Sad	72	0.78	73	0.76	0.23	0.21	1.00	0.08
		Angry	72	0.98	73	0.96	0.07	3.88	< 0.001	0.33
		Disgust	72	0.91	73	0.44	0.20	10.47	< 0.001	2.39
		Neutral	72	0.93	73	0.79	0.17	8.31	< 0.001	0.80
		Overall	72	0.83	73	0.71	0.07	10.23	< 0.001	1.62
	Overall		72	0.79	73	0.69	0.04	10.48	< 0.001	2.29

Note: The group comparisons between males and females were made using the *Wilcoxon-rank-sum* test. *SD(Δ)* = standard deviation of the difference of *relative frequencies female* and *relative frequencies male*. *Positive z-scores* indicate that the emotional portrayals had higher identification rates when spoken by female actors, whereas *negative z-scores* denote that the performance was higher when the emotions were spoken by male actors. The tests were conducted for each emotion separately, across all emotions and stimulus types. All *p-values* were Bonferroni corrected.

positive nouns ($\chi^2_{(5)}$ =17.94, p =.044). This second order interaction pattern is explained by the inspection of the average ratings showing different gender patterns conditional on emotion categories (see ***Figure 3***). No second order interactions were found across stimuli types [*Group Words* ($\chi^2_{(5)} = 15.00$, $p = .164$), *Group Sentences* ($\chi^2_{(6)} = 4.50$, $p = 1.00$)] and for any of the other stimuli sub-sets: *Anna* ($\chi^2_{(3)} = 2.57$, $p = 1.00$), *pseudo-words* ($\chi^2_{(5)} = 10.04$, $p = .927$), *semantic negative nouns* ($\chi^2_{(5)} = 6.63$, $p = 1.00$), *semantic neutral*

nouns ($\chi^2_{(5)} = 2.36$, $p = 1.00$), *affect bursts* ($\chi^2_{(6)} = 4.94$, $p = 1.00$), *pseudo-sentences* ($\chi^2_{(6)} = 5.64$, $p = 1.00$), *lexical sentences* ($\chi^2_{(6)} = 5.70$, $p = 1.00$) and *neutral sentences* ($\chi^2_{(5)} = 2.01$, $p = 1.00$).

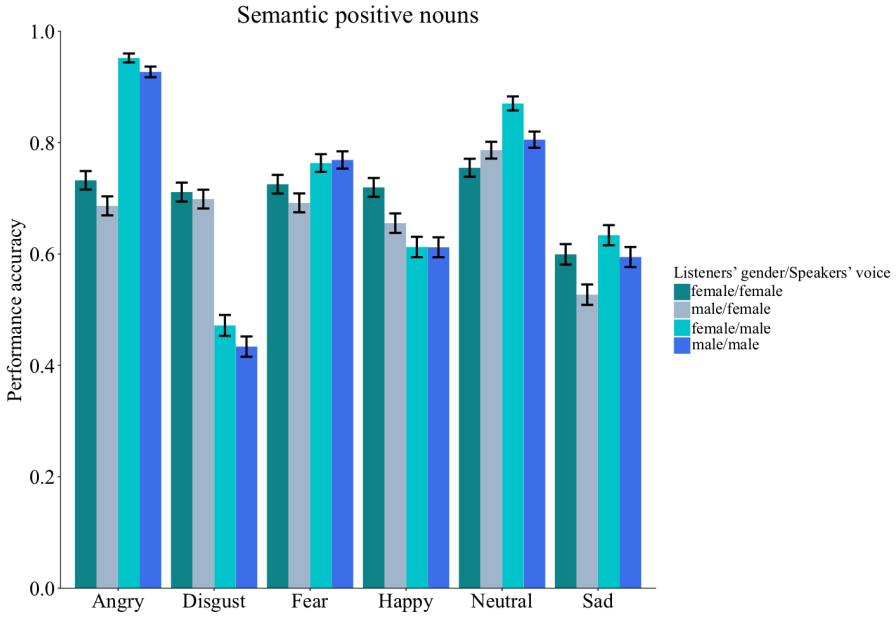


Figure 3| *Group Words* (*n* = 145, 71 females)
Bar charts showing the performance accuracy by listeners' gender, speakers' gender and emotion categories. Error bars represent the standard error. As it can be observed, the second order interaction pattern is explained by the inspection of the average ratings showing different gender patterns conditional on emotion categories. For instance, female listeners had higher recognition accuracy for the emotion category happiness encoded in the female voice and higher recognition accuracy for the neutral category encoded in the male voice. In contrast, male listeners had lower recognition accuracy for the emotion category sad when spoken by a female.

2.4 Discussion

The present study aimed at investigating gender differences in the recognition of vocal emotions. Specifically, we investigated any gender-specific advantage for the decoding of vocal emotions that were presented across a variety of stimulus types and emotion categories. A second objective was to assess whether the speakers' gender impacts on identification accuracy for different types of vocal emotions. Finally, we explored potential interactions between listeners' and speakers' gender for the identification of vocal emotions. The stimuli used in this study included a wide range of vocal utterances (e.g., words/pseudo-words, sentences/pseudo-sentences, affect bursts) that were expressing different emotions [i.e., anger, disgust, happy, fear, sadness and surprise, or no emotion (neutral)]. These characteristics of the stimulus set allowed us to assess gender differences for the recognition of vocal emotions in a differentiated manner and to provide parameter estimates on the magnitude of these effects. Especially the latter represents a largely neglected aspect within the vocal emotion literature.

Overall, our results showed that in each of the databases there were large differences in the recognition rates between emotions confirming well-established findings that recognition accuracy depends largely on the emotion category concerned (e.g., K. R. Scherer et al., 2001). Furthermore, we observed that performance accuracy is modulated by listeners' and speakers' gender and can significantly vary across stimulus types and emotion categories. Finally, we found that speaker gender had a significant impact on how listeners

judged specific emotions from the voice. These findings will be discussed in detail in the following sub-sections.

2.4.1 Performance accuracy by listeners' gender

We observed a significant main effect of gender reflecting that females outperformed males at categorizing emotions in vocal stimuli. The direction of this effect is consistent with previous findings on the recognition of non-verbal expressions of emotion (e.g., Collignon et al., 2010; Hall, 1978; Kret & De Gelder, 2012; A. E. Thompson & Voyer, 2014; Wingenbach et al., 2018) and emotional prosody in particular (e.g., Bonebright et al., 1996; Keshtiari & Kuhlmann, 2016; Paulmann & Uskul, 2014; K. R. Scherer et al., 2001).

An interesting pattern observed in our study is that females outperformed males when listening to emotionally produced pseudo-speech. Although the differences between females and males were not significant for emotion-specific categories the results clearly showed that overall females outperformed males when recognizing emotions from pseudo-sentences and had slightly higher recognition rates when decoding emotions from pseudo-words. As pseudo-speech lacks semantic meaning, one possible explanation for this effect is that women compared to men are better decoders under conditions of minimal stimulus information [see for example, *child-rearing hypothesis* (Babchuk et al., 1985) according to which females due to their role as primary caretakers have developed 'evolved adaptions' hypothesized to include the fast and accurate decoding especially, for negative emotions]. However, the female advantage was significant only for pseudo-sentences and, thus until further evidence is provided on the robustness of these effects, this interpretation should be approached with caution. One example related to this interpretation is a large-sample study ($N = 5872$) conducted by A. H. Fischer et al. (2018) that failed to replicate earlier findings assuming that females are better than males when categorizing discrete emotions in faces under situations of minimal stimulus information.

Previous studies revealed that females score higher than males in decoding specific emotions such as happiness, fear, sadness or neutral expressions and that both genders perform equally well for emotions spoken in an angry and disgusted tone of voice (for an overview see Table 1). Our results partially support these findings. On the one hand, we were able to show that the performance accuracy between females and males did not differ for emotions spoken in an angry and disgusted tone of voice. The absence of a gender specific advantage for decoding a socially salient vocal emotion such as anger may be because humans (and other primates) are biologically prepared or "hard-wired" (Öhman, 1993) to respond rapidly to specific stimuli (e.g., screams; alarm calls) in the environment, independent of gender. Moreover, it has been suggested that anger and disgust are expressions that signal the rejection of something or someone (Schirmer, 2013) and, thus, one could argue that they place an equal demand on attentional resources regardless of gender. On the other hand, we found no evidence that females outperform males when decoding distressing (i.e., sad, fear), happy and neutral emotions from the voice. Similar to other findings on gender differences in emotion recognition (e.g., A. H. Fischer et al., 2018), the magnitude between genders we observed for the decoding of vocal emotions was

relatively small. In our study, however, the direction of this effect consistently showed a female advantage for these specific emotions. Incorporating this pattern in biological and socialization models, one could assume that due to their ascribed nurturing, affiliative and less dominant role (Hess, Adams Jr, & Kleck, 2005; Schirmer, 2013, for a review) women might have developed a higher sensitivity to minimal affective signals (e.g., recognition of infants' fleeting and subtle emotional signals) which may contribute to their advantage in understanding other persons' emotional states. Nevertheless, the variety of methodologies used in previous research posits difficulties when aiming to draw a conclusion against or in favor of an 'female advantage' towards these specific emotions. Although our results partially suggest that females may have an advantage when decoding emotions from the voice, these effects might relate to the specific stimulus sets rather than female sensitivity towards particular emotions. It seems plausible that this finding is attributable to the different number of emotional categories included in the stimulus sets. For instance, some stimulus sets covered less emotional categories (e.g., *Anna*) than the options participants were offered to choose from. Offering emotional categories not included in the stimulus set, could lead to a systematic error in the face of a dichotomous choice between an emotion included in the set and one not included (e.g., happy versus surprise). Another possible explanation could be a bias towards 'negative' emotions, due to the majority of emotional categories being negative (four out of seven options). In addition, research has shown that speakers' pitch contour largely depends on the type and length of stimuli (e.g., words versus sentences; Jürgens et al., 2015, for a discussion) and, thus, one could argue that the results on emotion specific effects were affected by the acoustic properties of stimuli (e.g., pitch, timing, voice quality; see Banse & Scherer, 1996; Juslin & Laukka, 2003, for details), which might have varied between the stimulus sets.

2.4.2 Performance accuracy of identifying vocal emotions by speakers' gender

Despite observing some variability between genders in the expression of emotions for certain types of stimuli (e.g., *Anna* and *nouns* with a semantic positive and negative connotation were better identified when spoken by males), overall performance accuracy was significantly higher when females expressed the emotions. While findings for emotion-specific categories are pretty much inconsistent across studies, our results showed significantly higher identification rates for disgust when spoken by a female. However, for the other emotions categories, this pattern was less straightforward than one would expect. For instance, the identification rates for portrayals of anger were not consistently higher when spoken by a male. Likewise, happy, fearful or sad tone of voice were not invariably better identified when the speaker was a female (see Table 1 for an overview on previous findings). Enhanced identification of women compared to men's emotional expressions has been shown in both, facial (e.g., Hall, 1984) and vocal domains (e.g., Belin et al., 2008; K. R. Scherer et al., 2001). Previous reviews addressing gender-related patterns for the expression of emotions have suggested that these predispositions emerge as a result of various factors ranging from biologically innate traits, social norms and skills to situational

contexts (e.g., see Chaplin, 2015; A. H. Fischer & LaFrance, 2015, for an overview). While the overall female advantage in the expression of emotions is advocated across studies on non-verbal communication (Hall, 1984), less clear is the evidence on why females and males differ in how well they can express particular emotions.

In our study, the mixed pattern for emotion-specific effects indicates that in the vocal domain, the reliability of emotion judgments is not systematically influenced by encoders' gender and the related stereotypes of emotional expressivity. Prior studies suggested that encoders' success in the speech channel may vary with the standardized utterance used (Banse & Scherer, 1996; Juslin & Laukka, 2003). As our stimulus sets were standardized utterances selected from validated databases we cannot clearly comment on similarities within or differences between the stimulus sets that might explain the observed mixed-pattern of results. One can assume, however, that in each database the instructions given to encoders' when portraying the emotions were different. This might have increased the chance that encoders differentially produced high- and low-intensity variants of vocal expressions of emotion and, thus one could speculate that, independent of gender, stimuli with higher-intensity were better identified than those with low intensity (Banse & Scherer, 1996; Juslin & Laukka, 2003). Another potential explanation for these variations in performance accuracy is, that in all databases the emotional expressions were recorded in a controlled setting through professional and non-professional actors. They were thus not real-life emotional expressions. While the methods of emotion simulation offer high experimental control, the validity of prosodic stimuli derived from these measures is limited (K. R. Scherer, 1986) and may boost *recognition accuracy* (Sauter & Fischer, 2018). Previous studies found that speakers often portray stereotypes of emotions and might differ in the quality of their emotional portrayals (e.g., one speaker might be very good at portraying happiness but not fear, whereas another speaker's performance might show the opposite pattern; Banse & Scherer, 1996; K. R. Scherer, 1986). More recent studies complement this evidence by showing that speakers with less acting experience might encounter difficulties when asked, for instance, to emote in a language devoid of meaning (e.g., Paulmann, Furnes, Bøkenes, & Cozzolino, 2016). Similarly, past work has shown that emotion categories sharing the same dimension of valence (e.g., happiness and surprise) and arousal (e.g., anger and fear) are more likely to be confused (e.g., Banse & Scherer, 1996). Thus, it is plausible that enacted emotions, expressed in isolation (i.e., without situational context) and belonging to the same valence category, might have challenged not only encoders' but also listeners' performance accuracy, thereby leading to ambiguous results. Finally, one could argue that the observed patterns in our results with regard to the identification accuracy of particular emotions from speakers' voice might not only be related to above-mentioned characteristics of our selected databases (e.g., types of stimuli, speakers acting experience, context) but they may also be reflected in the similarities and differences of acoustic and spectral profiles of emotional inflections in spoken language and non-verbal vocalizations (see, Banse & Scherer, 1996; Juslin & Laukka, 2001; Sauter, Eisner, Calder, & Scott, 2010, for details), which can be independent of encoders' gender.

2.4.3 Interplay between listeners, speakers gender and emotion categories

In contrast to previous findings (e.g., Belin et al., 2008; Bonebright et al., 1996) an interesting pattern we observed in our study is related to the significant interaction between listeners' gender, speakers' gender and emotions for sematic positive nouns. This showed that females were more sensitive to happy expressions spoken by a female, while sensitivity increased for angry, neutral, disgust and sad expressions when spoken by a male. Although recognition accuracy seems to be contingent on the emotion being decoded as well as the speaker's gender, it is not clear whether the influence of encoder gender on these emotions reflects systematic properties of how these emotions are decoded and labelled, or whether certain artefacts may have been introduced by the semantic category of the stimuli (e.g., positive content spoken in an angry voice). As this pattern was present only for this type of stimuli we do not have a clear explanation for this effect. At most, one could speculate that females might be disposed to display fast and accurate decoding strategies in the face of an apparently conflicting message presented through semantics to detect credible cues about a speaker's true attitude and intentions. Words with a positive and negative semantic connotation, for instance, were found to have a processing advantage over neutral words (e.g., Schacht & Sommer, 2009a, 2009b) and, thus one may speculate that this type of stimuli (here meaningful nouns) that either express an emotional state (e.g., happiness) or elicit one (e.g., Satan) provoke differential responses in females and males.

2.4.4 Strengths, limitations and future research

As emphasized by previous research and corroborated by our data, there are several advantages to control for factors believed to be central when assessing emotion recognition ability. First, the ecological validity of emotion recognition tasks can be expected to increase when a large number of stimuli containing a wider range of emotional expressions is studied. Second, employing gender-balanced samples allows the control of possible main effects in emotion recognition ability while examining potential interaction effects between decoders' and encoders'. Finally, presenting participants with one out of several emotions reduces the likelihood of judges arriving at the correct answer by using exclusion and probability rules. Given that gender differences in the recognition of emotions are generally reported as small or even absent, the present study extends previous findings to show that the female advantage becomes more evident when using a variety of stimuli, a larger number of speakers and a wider range of emotions. Although, we agree to some extent with proponents of gender similarity hypothesis (e.g., Hyde, 2014) that this female advantage should not be over-interpreted, our results clearly indicate that in the vocal domain, there was an underlying consistency towards a female advantage across a wide range of presented stimuli. Therefore, we believe that before under-interpreting these effects, one should consider them within the larger context of the more recent literature (e.g., Wingenbach et al., 2018), which similar to our study, demonstrated that improved methodologies and analysis (e.g., balanced design) help to assess the differences between genders in a more representative and generalizable fashion.

In our study, results showed some strong differences favoring each gender when decoding specific emotions from speakers' voice yet, this pattern was less straightforward than we expected. Although all selected stimuli were from validated databases, the variations within our results may simply reflect inconsistent procedures attributable to database characteristics (e.g., speakers' training, baseline vocal qualities, recording conditions). Moreover, it should be noted that despite using a variety of stimuli the number of speakers for some stimulus types was quite small (e.g., pseudo-words; lexical sentences). This makes it hard to generalize the effects regarding speaker gender to other speech databases. Future research should, thus, control for these factors and, seek to replicate findings on gender differences in the recognition of vocal emotions by using datasets of stimuli that include fully naturalized speech in emotion-related states to further increase ecological validity.

The absence of certain emotional categories within the databases and the fixed alternatives of emotional categories listeners had to choose from, might have led to lower accuracy in performance due to higher levels of cognitive load imposed by the task format. We chose a fixed-choice response format to compare the results with the majority of prior literature. However, this format may be less ecologically valid (Russell, 1994) and thus, it has been suggested that tasks including "other emotion" as a response alternative (Frank & Stennett, 2001), visual analog scales (Young et al., 2017) or open-ended perspective taking (Cassels & Birch, 2014) may prove more sensitive when measuring individuals' ability to recognize emotions. Moreover, our experiment might have been affected by common method variance such as, assessment context (i.e., laboratory), item complexity (e.g., the perception of surprise might be interpreted as positive or negative) and mood state (for a comprehensive review, see Podsakoff, MacKenzie, Lee, & Podsakoff, 2003).

An unexpected finding within our study was the significant age difference between males and females in both groups. Although several studies demonstrated that advancing age is associated with lower accuracy performance in the recognition of vocal emotions (e.g., Lima et al., 2014; Paulmann & Kotz, 2008), our cohort was rather close in age (i.e., the older adults were not as old as populations reported in the literature). Thus, in future studies it would be interesting to clarify whether there is a critical earlier age period for emotional prosody recognition. This could be done, for example, by testing balanced groups of similar ages (e.g., 18-23; 24-29; 30-35) in order to specify a point of time at which emotional prosody recognition might start to decline with age.

Moreover, it has been suggested that prosodic acoustic parameters (e.g., speech melody, loudness), among other cues (e.g., semantics), provide listeners with a general understanding of the intended emotion and, thus, contribute in a cumulative fashion to the communication and recognition of emotions (W. F. Thompson & Balkwill, 2006, 2009). Future studies could explore how much of the variance in recognition rates is explained by similarities or differences in the acoustic attributes of emotive speech and assess the extent to which listeners use these acoustic parameters as a perceptual cue for identifying the portrayed emotion.

The present findings help to establish whether recognition accuracy differs according to listeners' and speakers' gender. Thus, an important step for future research will be

to evaluate theories regarding *why* these differences or similarities may occur by taking into account evolutionary, cognitive-learning, socio-cultural, and expectancy-value theories (Hyde, 2014).

Previous research suggested that the visual-modality conveys higher degrees of positivity-negativity, whereas the voice incorporates higher degrees of dominance-submission (e.g., Hall, 1984). Thus, one interesting line of future investigation could explore whether females specialize in visual and males in vocal communication. Finally, as the present study evidenced some differences in emotion decoding and encoding in the auditory modality, it would be worthwhile to investigate how these differences relate to audio-visual integration of emotional signals among men and women. The combination of recognition data with physiological measures (e.g., peripheral indicators of emotional responses), psychosocial (e.g., personality traits) and demographic variables (e.g., age, education), as well as, self-reported trait measures of emotional intelligence and tests to assess participants' ability for sustained attention during an experiment, could help to assess gender differences in emotion recognition in an even more differentiated manner.

2.4.5 Conclusion

The present study replicates earlier research findings while controlling for several previously unaddressed confounds. It adds to the literature on gender differences for the recognition of vocal emotions by showing a female advantage in decoding accuracy and by establishing that females' emotional expressions are more accurately identified than those expressed by men. Results explain inconsistencies in the past literature in which findings of female superiority for identifying vocal emotions remain mixed by highlighting that the effect emerges for particular stimulus categories and under controlled environments. The partially mixed pattern of results in the current experimental task should be further investigated in natural settings, to assess whether males and females are attuned towards specific emotions in more realistic contexts.

Funding

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project number 254142454/GRK 2070.

Acknowledgements

The authors thank Annika Grass and Florian Niefind for technical support, and Kurt Hammerschmidt and Lars Penke for fruitful discussions. We would further like to thank the editor and reviewers for their helpful comments on earlier versions of our manuscript.

2.5 Supplementary Material

Table S1 (a) | *Global models across all stimuli types for* ***Group Words***

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			82341	103392	
Age	1	3.90	82340	103328	0.838
Confidence	2	4650.90	82338	98738	< .001
Speaker gender	1	42.30	82337	98695	< .001
Participants gender	1	51.70	82336	98644	< .001
Emotions	5	4853.80	82331	93790	< .001
Stimuli types	4	826.20	82327	92964	< .001
Participants	142	2866.10	82185	90097	< .001
Speaker gender x Participants gender	1	0.30	82184	90097	1.00
Age x Stimuli types	4	12.90	82180	90084	0.191
Confidence x Stimuli types	8	89.70	82172	89995	< .001
Speaker gender x Stimuli types	4	142.80	82168	89852	< .001
Participants gender x Stimuli types	4	16.40	82164	89835	0.038
Emotions x Stimuli types	18	1097.80	82146	88738	< .001
Age x Speaker gender	1	9.70	82145	88728	0.030
Age x Emotions	5	78.20	82140	88650	< .001
Participants gender x Emotions	5	26.60	82135	88623	< .001
Speaker gender x Emotions	5	842.30	82130	87781	< .001
Confidence x Participants	286	1010.60	81844	86770	< .001
Speaker gender x Participants gender x Emotions	5	15.00	81839	86755	0.164

Resid. Df = residual degrees of freedom; *Resid. Dev.* = Residual deviance. *P-values* were Bonferroni corrected for multiple testing.

Table S1 (b) | *Global models across all stimuli types for* ***Group Sentences***

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			68124	78330	
Age	1	13.20	68123	78317	0.003
Confidence	2	5124.9	68121	73192	< .001
Speaker gender	1	589.4	68120	72603	< .001
Participants gender	1	5.2	68119	72597	0.332
Emotions	6	6956.0	68113	65641	< .001
Stimuli types	3	693.8	68110	64948	< .001
Participants	142	1344.3	67968	63603	< .001
Speaker gender x Participants gender	1	0.1	67967	63603	1.00
Age x Stimuli types	3	1.0	67964	63602	1.00
Confidence x Stimuli types	6	149.2	67958	63453	< .001
Speaker gender x Stimuli types	3	18.5	67955	63435	0.003
Participants gender x Stimuli types	3	22.8	67952	63412	< .001
Emotions x Stimuli types	17	1990.4	67935	61421	< .001
Age x Speaker gender	1	3.7	67934	61418	0.826
Age x Emotions	6	23.4	67928	61394	0.005
Participants gender x Emotions	6	19.6	67922	61375	0.029
Speaker gender x Emotions	6	726.7	67916	60648	< .001
Confidence x Participants	285	517.3	67631	60130	< .001
Speaker gender x Participants gender x Emotions	6	4.5	67625	60126	1.00

Resid. Df = residual degrees of freedom; *Resid. Dev.* = Residual deviance. *P-values* were Bonferroni corrected for multiple testing.

Table S2 (a) | *Quasi-binomial logistic model for* ***Anna***

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			12756	16889	
Age	1	2.97	12755	16886	1.00
Confidence	2	1094.62	12753	15792	< .001
Speaker gender	1	75.13	12752	15717	< .001
Participants gender	1	0.03	12751	15717	1.00
Emotions	3	2463.87	12748	13253	< .001
Participants	142	498.30	12606	12754	< .001
Speaker gender x Participants gender	1	1.41	12605	12753	1.00
Age x Speaker gender	1	0.43	12604	12753	1.00
Age x Emotions	3	1.68	12601	12751	1.00
Participants gender x Emotion	3	7.61	12598	12743	0.644
Speaker gender x Emotion	3	211.41	12595	12532	< .001
Confidence x Participants	278	364.91	12317	12167	< .001
Speaker gender x Participants gender x Emotion	3	2.57	12314	12164	1.00

Resid. Df = residual degrees of freedom; *Resid. Dev.* = Residual deviance. *P-values* were Bonferroni corrected for multiple testing.

Table S2 (b) | *Quasi-binomial logistic model for* ***Pseudo-words***

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			17394	22420	
Age	1	8.57	17393	22411	0.042
Confidence	2	959.64	17391	21452	< .001
Speaker gender	1	22.26	17390	21429	< .001
Participants gender	1	29.18	17389	21400	< .001
Emotions	5	1060.19	17384	20340	< .001
Participants	142	1072.58	17242	19267	< .001
Speaker gender x Participants gender	1	0.85	17241	19267	1.00
Age x Speaker gender	1	1.09	17240	19266	1.00
Age x Emotions	5	14.83	17235	19251	0.136
Participants gender x Emotion	5	15.18	17230	19236	0.117
Speaker gender x Emotion	5	202.22	17225	19033	< .001
Confidence x Participants	275	474.52	16950	18559	< .001
Speaker gender x Participants gender x Emotion	5	10.04	16945	18549	0.927

Resid. Df = residual degrees of freedom; *Resid. Dev.* = Residual deviance. *P-values* were Bonferroni corrected for multiple testing.

Table S2 (c) | *Quasi-binomial logistic model for* ***Semantic positive nouns***

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			17396	21343	
Age	1	0.15	17395	21343	1.00
Confidence	2	816.03	17393	20527	< .001
Speaker gender	1	0.43	17392	20527	1.00
Participants gender	1	20.14	17391	20506	< .001
Emotions	5	616.96	17386	19890	< .001
Participants	142	795.30	17244	19094	< .001
Speaker gender x Participants gender	1	0.06	17243	19094	1.00
Age x Speaker gender	1	4.67	17242	19090	0.414
Age x Emotions	5	39.31	17237	19050	< .001
Participants gender x Emotion	5	6.73	17232	19044	1.00
Speaker gender x Emotion	5	462.14	17227	18581	< .001
Confidence x Participants	280	421.42	16947	18160	< .001
Speaker gender x Participants gender x Emotion	5	17.94	16942	18142	0.044

Resid. Df = residual degrees of freedom; *Resid. Dev.* = Residual deviance. *P-values* were Bonferroni corrected for multiple testing.

Table S2 (d) | *Quasi-binomial logistic model for* ***Semantic negative nouns***

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			17395	21225	
Age	1	5.24	17394	21219	0.274
Confidence	2	970.85	17392	20249	< .001
Speaker gender	1	71.74	17391	20177	< .001
Participants gender	1	8.38	17390	20168	0.046
Emotions	5	735.54	17385	19433	< .001
Participants	142	815.60	17243	18617	< .001
Speaker gender x Participants gender	1	0.25	17242	18617	1.00
Age x Speaker gender	1	2.07	17241	18615	1.00
Age x Emotions	5	21.35	17236	18594	0.008
Participants gender x Emotion	5	14.96	17231	18579	0.124
Speaker gender x Emotion	5	280.14	17226	18298	< .001
Confidence x Participants	280	496.20	16946	17802	< .001
Speaker gender x Participants gender x Emotion	5	6.63	16941	17796	1.00

Resid. Df = residual degrees of freedom; *Resid. Dev.* = Residual deviance. *P-values* were Bonferroni corrected for multiple testing.

Table S2 (e) | *Quasi-binomial logistic model for* ***Semantic neutral nouns***

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			17396	21190	
Age	1	1.07	17395	21188	1.00
Confidence	2	888.61	17393	20300	< .001
Speaker gender	1	3.05	17392	20297	0.997
Participants gender	1	9.13	17391	20288	0.029
Emotions	5	1603.56	17386	18684	< .001
Participants	142	854.44	17244	17830	< .001
Speaker gender x Participants gender	1	2.80	17243	17827	1.00
Age x Speaker gender	1	8.62	17242	17818	0.038
Age x Emotions	5	35.03	17237	17783	< .001
Participants gender x Emotion	5	3.12	17232	17780	1.00
Speaker gender x Emotion	5	465.36	17227	17315	< .001
Confidence x Participants	275	476.61	16952	16838	< .001
Speaker gender x Participants gender x Emotion	5	2.36	16947	16836	1.00

Resid. Df = residual degrees of freedom; *Resid. Dev.* = Residual deviance. *P-values* were Bonferroni corrected for multiple testing.

Table S2 (f) | *Quasi-binomial logistic model for* ***Affect bursts***

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			10146	10059.1	
Age	1	0.87	10145	10058.2	1.00
Confidence	2	1116.13	10143	8942.1	< .001
Speaker gender	1	40.24	10142	8901.8	< .001
Participants gender	1	2.83	10141	8899.0	0.698
Emotions	6	1113.20	10135	7785.8	< .001
Participants	142	303.77	9993	7482.0	< .001
Speaker gender x Participants gender	1	0.21	9992	7481.8	1.00
Age x Speaker gender	1	1.38	9991	7480.4	1.00
Age x Emotions	6	14.71	9985	7465.7	0.047
Participants gender x Emotion	6	5.90	9979	7459.8	1.00
Speaker gender x Emotion	6	243.28	9973	7216.5	< .001
Confidence x Participants	278	362.30	9695	6854.2	< .001
Speaker gender x Participants gender x Emotion	6	4.94	9689	6849.3	1.00

Resid. Df = residual degrees of freedom; *Resid. Dev.* = Residual deviance. *P-values* were Bonferroni corrected for multiple testing.

Table S2 (g) | *Quasi-binomial logistic model for* ***Pseudo-sentences***

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			20292	26212	
Age	1	2.38	20291	26209	1.00
Confidence	2	1318.73	20289	24890	< .001
Speaker gender	1	173.65	20288	24717	< .001
Participants gender	1	7.44	20287	24709	0.068
Emotions	6	2784.06	20281	21925	< .001
Participants	142	524.65	20139	21401	< .001
Speaker gender x Participants gender	1	0.31	20138	21400	1.00
Age x Speaker gender	1	7.21	20137	21393	0.078
Age x Emotions	6	4.62	20131	21389	1.00
Participants gender x Emotion	6	17.36	20125	21371	0.075
Speaker gender x Emotion	6	1276.99	20119	20094	< .001
Confidence x Participants	278	347.52	19841	19747	0.005
Speaker gender x Participants gender x Emotion	5	5.64	19835	19741	1.00

Resid. Df = residual degrees of freedom; *Resid. Dev.* = Residual deviance. *P-values* were Bonferroni corrected for multiple testing.

Table S2 (h) | *Quasi-binomial logistic model for* ***Lexical sentences***

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			20291	22122	
Age	1	5.80	20290	22116	0.171
Confidence	2	773.30	20288	21343	< .001
Speaker gender	1	154.70	20287	21188	< .001
Participants gender	1	5.50	20286	21183	0.211
Emotions	6	3745.60	20280	17437	< .001
Participants	142	654.00	20138	16783	< .001
Speaker gender x Participants gender	1	0.20	20137	16783	1.00
Age x Speaker gender	1	1.50	20136	16782	1.00
Age x Emotions	6	37.50	20130	16744	< .001
Participants gender x Emotion	6	9.60	20124	16735	1.00
Speaker gender x Emotion	6	194.50	20118	16540	< .001
Confidence x Participants	279	506.70	19839	16033	< .001
Speaker gender x Participants gender x Emotion	6	5.70	19833	16028	1.00

Resid. Df = residual degrees of freedom; *Resid. Dev.* = Residual deviance. *P-values* were Bonferroni corrected for multiple testing.

Table S2 (i) | *Quasi-binomial logistic model for* ***Neutral sentences***

Model terms	***Df***	***Deviance***	***Resid. Df***	***Resid. Dev***	***Pr(>Chi)***
NULL			17397	20973	
Age	1	0.08	17396	20973	1.00
Confidence	2	899.54	17394	20074	< .001
Speaker gender	1	0.93	17393	20073	1.00
Participants gender	1	5.35	17392	20067	0.243
Emotions	5	1332.93	17387	18734	< .001
Participants	142	879.44	17245	17855	< .001
Speaker gender x Participants gender	1	1.67	17244	17853	1.00
Age x Speaker gender	1	10.64	17243	17843	0.011
Age x Emotions	5	37.62	17238	17805	< .001
Participants gender x Emotion	5	4.05	17233	17801	1.00
Speaker gender x Emotion	5	449.41	17228	17352	< .001
Confidence x Participants	275	494.17	16953	16857	< .001
Speaker gender x Participants gender x Emotion	5	2.01	16948	16855	1.00

Resid. Df = residual degrees of freedom; *Resid. Dev.* = Residual deviance. *P-values* were Bonferroni corrected for multiple testing.

Chapter 3

Emotion Recognition and Confidence Ratings Predicted by Vocal Stimulus Type and Acoustic Parameters

Abstract

Our speech expresses emotional meaning, not only through words, but also through certain attributes of our voice, such as pitch or loudness. These prosodic attributes are well-documented within the vocal emotion literature. However, there is considerable variability in the types of stimuli and procedures used to examine their influence on emotion recognition. In addition, the confidence we have in our assessments of another person's emotional state has been argued to strongly influence performance accuracy in emotion recognition tasks. Nevertheless, such associations have rarely been studied previously. We addressed this knowledge gap by examining the impact of vocal stimulus type and prosodic speech attributes on emotion recognition and a person's confidence in a given response. We analyzed a total of 1038 emotional expressions spoken in an *angry, disgusted, fearful, happy, neutral, sad* and *surprised* tone of voice according to a baseline set of prosodic acoustic parameters ($N = 13$). Two classification procedures (*linear discriminant analysis* and *random forest*) established that these acoustic measures provided sufficient discrimination between expressions of emotional categories to permit accurate statistical classification. Logistic regression- and linear models showed that emotion recognition and confidence judgments essentially depended on stimulus material as they could be predicted by different constellations of acoustic features. Results also demonstrated that emotional expressions which were correctly identified elicited confident judgments. Together, these findings extend previous work by showing that vocal stimulus type and prosodic attributes of speech strongly influence emotion recognition and listeners' confidence in a given response.[1]

Keywords: Vocal Emotion Recognition, Confidence Judgements, Acoustic Parameters, Speech and Non-speech Stimuli, Classification Methods

[1] Lausen, A., Hammerschmidt, K., & Schacht, A. (2019). Emotion recognition and confidence ratings predicted by vocal stimulus type and acoustic parameters. *doi: 10.31234/osf.io/kqy2n*

3.1 Introduction

The ability to correctly understand and appropriately respond to emotions, plays an important role in everyday social interactions (e.g., Chronaki, Wigelsworth, Pell, & Kotz, 2018; Juslin & Scherer, 2005). In verbal communication, for instance, humans do not merely consider *what* their interlocutors are saying (i.e., semantic meaning), but also *how* they are conveying the spoken information (e.g., high/low pitch of their voice). An all-encompassing term for such vocal qualities of speech is *prosody*. Although in social interactions, the emotional expression of a message is usually conveyed by various channels (i.e., voice, face, body), it has been demonstrated that prosody supports correct interpretations of utterances (Paulmann, 2016; W. F. Thompson & Balkwill, 2009), independently of linguistic comprehension (Kitayama & Ishii, 2002). Another factor argued to influence the correct recognition and interpretation of emotions is *metacognition*, i.e. the capacity to actively monitor and reflect upon one's own performance (Dunlosky & Metcalfe, 2008). Accurate metacognition was argued to promote correct and confident interpretations, while uncertainty in the interpretation of ambiguous emotional expressions may prompt the perceiver to seek additional information until a confident assessment can be made (Kelly & Metcalfe, 2011). To better understand the mechanisms underlying the recognition of emotions from the voice, the present study examined how different types of vocal stimuli and their acoustic attributes influence listeners' recognition of emotions and confidence ratings.

In their endeavor to assess the recognition of emotions from prosody, researchers created a wide variety of stimulus materials. Some decided to use sentences as stimuli because they have been argued to have higher ecological validity (Sauter, 2006). However, as emotions are not expressed to the same degree in each word of a sentence it has been suggested that such long-lasting stimuli might contain increased variation and noise in the signal (Sauter, 2006). Thus, other investigators choose to use single words as stimulus material as they do not "dilute" the characteristics of a specific emotion (K. Hammerschmidt & Jürgens, 2007). Several studies reported specific emotions to be more easily recognized if semantic information is available even if semantics are irrelevant to the given task (e.g., Ben-David, Multani, Shakuf, Rudzicz, & van Lieshout, 2016; Paulmann et al., 2016), while others reported that semantic information might facilitate or interfere with a listener's judgment about the emotional content of the stimulus when spoken in a congruent or incongruent prosody (e.g., Kotz & Paulmann, 2007; Mitchell, Elliott, Barry, Cruttenden, & Woodruff, 2003; Nygaard & Lunders, 2002; Pell & Kotz, 2011). To ensure that semantic information does not confound prosody of the spoken stimuli, previous research used the speech-embedded material in a pseudo-language (i.e., an artificially created language devoid of meaning). This approach represents a useful way to neutralize or mask the semantic content while retaining the prosodic information (Banse & Scherer, 1996; Rigoulot et al., 2013). Studies on the identification of vocal emotions from pseudo-utterances found overall recognition rates for discrete emotions to be significantly higher than chance (e.g., Pell, Paulmann, et al., 2009; K. R. Scherer et al., 2001). However, the analysis of emotional prosody in isolation (i.e., without lexico-semantic content) might not only increase the artifice of the acted emotions but could also lead to poorer decoding accuracy (Parsons, Young, Craske,

Stein, & Kringelbach, 2014). Thus, it has been suggested that non-speech sounds or affect bursts (e.g., laughter, screams) are the only reliable type of stimuli comprising the most 'natural and ancient language of emotion communication whose expressiveness no words can ever achieve' (Bostanov & Kotchoubey, 2004, p.259). Indeed, several studies demonstrated that non-speech sounds provide more discriminable emotion information than do speech-embedded prosodic signals (Hawk et al., 2009; Pell et al., 2015) and are effective in eliciting the attention of others (K. R. Scherer, 1994), especially, in a potential threat life-situation (e.g., hearing someone's scream).

The discussion whether the recognition of vocal emotions from speech-embedded materials has an advantage over non-speech embedded materials or vice-versa is far from settled. The extraction of acoustic cues from their created materials led, however, to an agreement among researchers that voice melody (rising/falling pitch), loudness, tempo and quality (stressed/breathy) are the most relevant paralinguistic features that speakers employ when expressing emotions (e.g., Goudbeek & Scherer, 2010). A series of statistics on these paralinguistic features has revealed that pitch or *fundamental frequency (F0)* related parameters (e.g., minimum, maximum, mean, jitter), *energy/amplitude*- (e.g., loudness, shimmer), *temporal*- (e.g., duration) and *quality* parameters (e.g., harmonics-to-noise ratio [HNR]) are amongst the most important 'candidates' for prosodic correlates of emotion in speech (e.g., Johnstone & Scherer, 2000; Juslin & Laukka, 2003). In their seminal work on the acoustic profiles of vocal emotion expression, Banse and Scherer (1996) examined paralinguistic features and their related acoustic parameters ($n = 29$) in 224 pseudo-utterances, each spoken in fourteen different emotions. The authors found that the emotion of the utterance predicted a large proportion of the variance in most of the acoustic variables, especially for mean F0 (50%) and mean energy (55%). Regressing participants' emotion ratings for each stimulus class, they found that this perceptual measure could be significantly predicted by a linear combination of a set of seven acoustic parameters (multiple correlation coefficient ranged between .16 for cold anger and .49 for sad). Implementing linear discriminant analysis (LDA) with jackknife and cross-validation procedures for the evaluation of classification errors they further showed that the overall patterns when categorizing emotions were similar to those of listeners' accuracy (LDA = 40% jackknife estimate of accuracy; LDA = 25% cross-validation estimate of accuracy; listeners = 48% accuracy). Subsequent studies conducted with different stimulus types (e.g., words, lexical and neutral sentences, affect bursts), with less or larger acoustic parameter sets (from 3 to 40 prosodic features), with the same (i.e., LDA) or other classification methods (e.g., k-nearest-neighbor classifier, random forest) obtained comparable results showing that both, classifiers and listeners perform similarly well when predicting emotion category membership based on the acoustic profiles of their utterances (e.g., K. Hammerschmidt & Jürgens, 2007; Juslin & Laukka, 2001; Noroozi, Sapiński, Kamińska, & Anbarjafari, 2017; Paulmann et al., 2008; Pichora-Fuller, Dupuis, & Van Lieshout, 2016; Sauter et al., 2010; Sbattella et al., 2014; Toivanen, Väyrynen, & Seppänen, 2004).

Together these findings allow the conclusion that prosodic acoustic parameters (among other cues, e.g., semantics) provide listeners with a general understanding of the intended

emotion and, thus, contribute in a cumulative fashion to the communication and recognition of emotions (e.g., W. F. Thompson & Balkwill, 2009). Nevertheless, it has been argued that using different types of stimuli, different sets of acoustic parameters and implementing various classification methods causes serious difficulties when interpreting the results across studies, endangering the accumulation of empirical evidence. Thus, adopting a baseline set of acoustic parameters and systematically analyzing their influence on emotion recognition ability within the various types of vocal stimulus material would improve methodological rigor and increase the reliability of findings (Bąk, 2016; Eyben et al., 2016; Juslin & Laukka, 2003).

As the subjective character of emotion recognition dictates a great variability in the way individuals interpret emotional messages, it has been argued that metacognition (i.e., the awareness of one's own knowledge) might impact judgments of accuracy in emotion recognition tasks. Kelly and Metcalfe (2011), for instance, investigated whether individuals can accurately predict and assess their performance on two face emotion recognition tasks (i.e., *Mind in the Eyes task* and *Ekman Emotional Expression Multimorph Task*). For each emotional expression, participants were asked to predict (1) their future performance in correctly identifying the emotions (i.e., prospective judgements) and (2) the accuracy regarding their confidence in the given responses (retrospective judgements). Results from the *Mind in the Eyes task* showed significantly higher scores for retrospective than prospective confidence judgements, however, no significant relationship between these judgements and performance accuracy was found. Even though in the *Emotional Expression Multimorph Task*, the gamma correlations were slightly greater for retrospective ($r = .43$) than prospective judgements ($r = .32$), the authors found a significant relationship between both types of judgements and performance accuracy. Based on these findings, they concluded that individuals who perform better in emotion recognition tasks are also more accurate in their metacognitive assessments. While some studies examining the perceptual-acoustic correlates of vocal confidence, showed that for listeners both linguistic and acoustic-prosodic cues are fundamental when making retrospective judgements about speakers' mental states (e.g., Jiang & Pell, 2014, 2017; Kimble & Seidel, 1991; K. R. Scherer, London, & Wolf, 1973), other studies demonstrated that in tasks assessing vocal expressions of emotion, listeners' confidence increased with stimulus duration (Pell & Kotz, 2011). For instance, Rigoulot et al. (2013) investigated the time course of vocal emotion recognition employing a modified version of an auditory gating paradigm. Results showed that, independent of stimulus presentation (forward or backward), listeners' confidence in categorizing the emotions increased significantly with longer gate intervals (i.e., number of syllables). This pattern of results clearly indicates that when assessing the recognition of vocal emotions, duration, among other acoustic parameters (e.g., pitch, loudness), progressively activates emotion-specific knowledge leading to higher accuracy and confidence ratings. While these findings reveal how much information is needed for listeners to consciously reflect on and categorize vocally-expressed emotions from paralinguistic attributes of speech, there is a lack of direct evidence examining the influence of vocal stimulus type and their related acoustic parameters on emotion recognition and confidence ratings.

Although there has been much research on the discrimination of emotions from speech, from the viewpoint of both human listeners and classification algorithms, comparing the results across studies is not a straightforward matter as performance accuracy was argued to essentially depend on the stimulus material and the extracted set of acoustic parameters. By implementing a standard set of acoustic parameters as baseline and two classification methods (i.e., linear discriminant classifier and random forest), the first aim of the present study was to investigate the extent listeners and classifiers use these acoustic parameters as a cue for identifying the portrayed emotion (i.e., by examining how much of the variance in recognition rates is explained by the acoustic attributes of emotive speech).

Research on metacognition has related this skill to vital aspects of socioemotional processes showing that confidence judgments are more accurate when given after than before a response to an emotion recognition task. New endeavors, however, are needed to document the extent emotion recognition and confidence ratings are predicted by the vocal stimulus type and their related acoustic parameters. These data would allow a more differentiated assessment of the factors assumed to impact both emotion recognition ability and a person's confidence in a given response. Thus, a secondary aim of this study was to examine whether and how listeners' performance accuracy and confidence judgements differ between certain types of vocal stimuli and specific emotion categories. A final aim was to assess whether retrospective confidence judgements are predicted by the correct recognition of vocal emotional expressions.

3.2 Method

The research project has been approved by the Ethical Committee of the *Georg-Elias-Mueller-Institute of Psychology*, University of Goettingen, Germany (*number 149*) and conducted in accordance with the ethical principles formulated in the *Declaration of Helsinki* (2013).

3.2.1 Participants

Two-hundred ninety participants (143 females, 147 males; *age range* = 18–36 years) completed the study after responding to advertisements posted on social media (e.g., Facebook) or to flyers distributed across the university campus. Participants averaged 23.83 years in age (SD = 3.73) with 62% having completed a general qualification for university entrance, 25% a bachelor degree, 12% a master degree and 1% a general certificate of secondary education. To reduce the length of the experiment, participants were allocated to two groups of equal size. One group listened to words and pseudo-words (*Group Words*, n = 145, M_{age} = 24.00, SD_{age} = 3.67), while the other group listened to affect bursts, sentences and pseudo-sentences (*Group Sentences*, n = 145, M_{age} = 23.66, SD_{age} = 3.80). No significant age difference between the two groups was observed ($t_{(288)}$= 0.786; p = 0.432; $CI_{95\%}$ = [-0.52; 1.21]). All participants were native speakers of German and reported no hearing difficulties.

3.2.2 Stimulus material & Acoustic analyses

One thousand thirty-eight emotional expressions spoken in an *angry*, *disgusted*, *fearful*, *happy*, *neutral*, *sad* and *surprised* tone of voice were sampled from established speech corpora or from researchers that developed their own stimulus materials (see Lausen & Schacht, 2018, for details). The stimulus material was analyzed for **frequency** related parameters (*mean fundamental frequency (F_0)*, *minimum F_0*, *maximum F_0*, *standard deviation F_0*, *jitter*), **energy/amplitude** related parameters (*shimmer*, *amplitude* [dB], *peak amplitude*, *mean HNR*, *maximum HNR*, *standard deviation HNR*) and **temporal features** (*duration*, *peak time*) using *GSU Praat Tools* script packages developed by Owren (2008), which allows batch processing during measurement (for details on the processing of acoustic parameters see *supplementary material*). Following the procedures of Goudbeek and Scherer (2010), Sauter et al. (2010) and Juslin and Laukka (2001) the measurements were made over the entire utterances, across all speakers and all items of the same type of stimulus.

A *linear discriminant analysis* (LDA) was then performed for each type of stimulus separately and across all stimuli in both groups to determine the optimal combination of the 13 above-mentioned acoustic parameters for predicting emotion category membership. In the analysis, acoustic measurements served as independent variables while the dependent variable was the intended emotional category. As the set of acoustic parameters was not very large, no feature selection method (e.g., stepwise analysis) was used to reduce the number of parameters. LDA is optimal if the acoustic parameters have a multivariate normal distribution with different means for each emotion and identical variance matrices for all emotions. However, if the underlying multivariate structure is more complex, other classification algorithms have been suggested to yield better performance (James, Witten, Hastie, & Tibshirani, 2013). To assess whether our LDA model shows better predictive performance than other classification techniques we implemented *random forest* (RF) as an additional classification algorithm. This ensemble classification methodology, which combines a large number of decision trees using different sets of predictors at each node of the trees, was argued to be a more robust alternative to discriminant analysis or multinomial regression as it allows a selection of the important potential predictors among a large number of variables with complex interactions (Anikin & Lima, 2018; Breiman, 2001, for a detailed explanation on RF see *supplementary material*). The two classification methods were compared by the estimated classification errors using 10-fold cross-validation.

3.2.3 Procedure & Experimental task

Up to four participants were invited to each experimental session, which lasted approximately 60 minutes. At arrival, the experimenter debriefed the participants about the aim of the study, i.e., to validate a set of auditory stimuli with emotional content. Prior to formal testing, participants signed a consent form and completed a short demographic questionnaire concerning age, gender and education level. Participants were informed that all stimuli would be presented only once, the number of presented emotions might vary from the number of categories given as possible choices, and some of the stimuli were

not supposed to carry any semantic meaning and might sound 'foreign'. After these instructions and completion of ten practice trials, participants started the main experiment, presented via *Presentation* software (Version 14.1, Neurobehavioral Systems Inc., Albany, CA). Stimuli were presented to the participants binaurally with *Bayerdynamic DT 770 PRO* headphones plugged-in in the tower box of a *Dell OptiPlex 780 SFF Desktop PC Computer*. To ensure equal physical volume of stimulus presentation across the four PCs, we measured the sound level meters of the ten practice stimuli with a professional sound level meter, *Nor140* (Norsonic, 2010, Lierskogen, Norway). No significant difference in volume intensity was observed ($F_{(3,27)} < 1$). Following each stimulus presentation listeners rendered two judgments: First, they classified which emotion was being expressed by the speaker from a list of seven categories presented on the computer screen. To assess metacognition, this rating was followed by a 7-point rating scale on the screen to estimate their confidence in the preceding response (*1 = not at all confident*; *7 = extremely confident*). ***Figure 1*** displays the course of the forced-choice task. The set of stimuli in *Group*

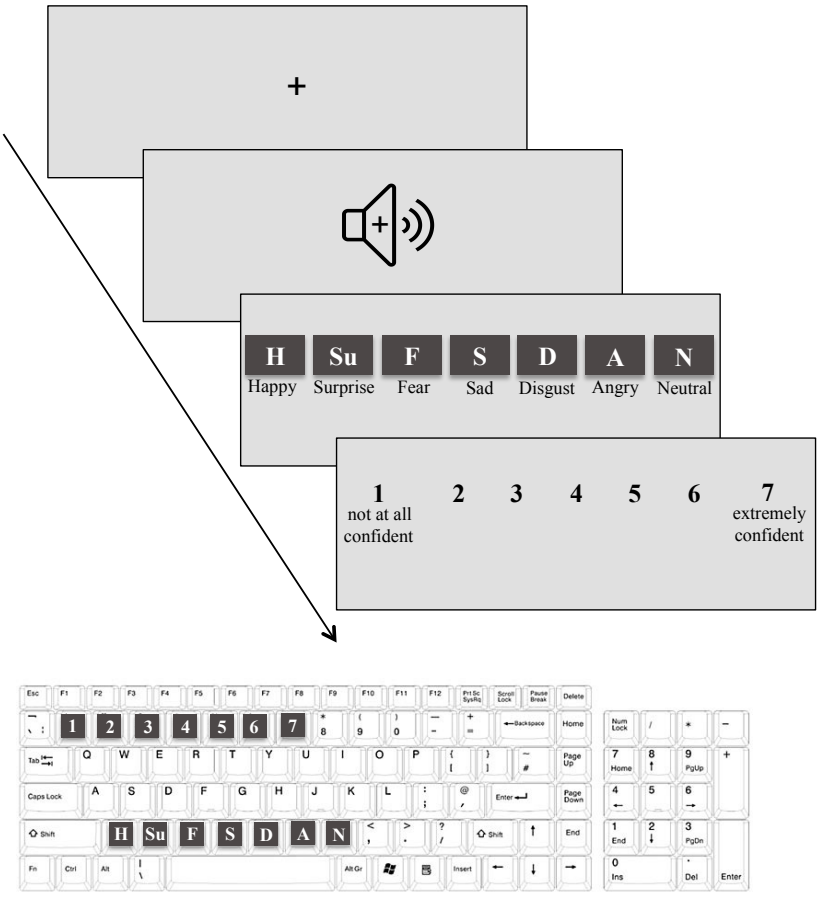


Figure 1 |*The course of the forced-choice task*
Each trial began with a fixation-cross appearing at the center of the screen, at which the participants were asked to fixate throughout the trial. The presentation of the stimuli was initiated by pressing the *Enter*-key. The auditory stimulus was then presented alongside the fixation cross. After the presentation of the stimulus the emotions panel with 7 categories (H = *Happy*; Su = *Surprise*; F = *Fear*; S = *Sad*; A = *Angry*; N = *Neutral*) followed by the confidence ratings panel were presented. Participants could hear the stimulus only once. The responses were made using the marked keyboard (*Z* to *M* for the emotion judgments, which were labeled corresponding to the emotion categories, and *1* to *7* for confidence ratings). There was no time limit for emotion judgments or confidence ratings. At the end of each block a visual message in the center of the screen instructed participants to take a break if they wished to or to press the *Spacebar* to proceed with the next block.

Words was split into three blocks (i.e., *Anna*, *Pseudo-words*, and *Nouns*), while in *Group Sentences* the set of stimuli was divided into four blocks (i.e., *Pseudo-sentences*, *Lexical Sentences*, *Neutral sentences*, and *Affect bursts*). The order of blocks and of the stimuli within each block were randomized. Blocks were separated by a break of self-determined duration. The reimbursement of participants consisted of 8€ or course credit.

3.2.4 Study design & Power analysis

To assess listeners' judgements of emotions and confidence in their judgements a *within-subjects design* was fitted for *Group Words* and *Group Sentences*. The design was balanced for emotion categories in each stimulus type. Independent within-subject factors were *stimuli types*, *acoustic parameters* and *emotion categories*. Dependent variables were *emotion recognition* and *confidence ratings*.

To assess whether we had enough power to answer our research questions, an *approximate correlation power analysis* was calculated and Bonferroni corrected for the 13 acoustic parameters. A sample size of 145 participants per group with a minimum set of stimuli per participant (70) allowed us to detect correlations of $r = 0.037$ with a type I error rate of 5% and power 80%. To describe the power to detect differences between emotion categories and stimulus types an *approximate Tukey's multiple pairwise comparisons power analysis* was computed. Assuming a minimum set of 10 stimuli for each emotion category and a sample size of 145 participants per group allowed us to detect a difference of 0.044 for recognition probability at 0.80 with a type I error rate of 5% and power 80%.

3.2.5 Statistical analysis

The data was analyzed by *generalized linear models* (*quasi-binomial logistic regression*) for the binary response variable emotion recognition and by linear models for the response variable confidence ratings. To find a reduced model that best explains the data on the 13 acoustic parameters for the two dependent variables a *backward stepwise variable selection* (R function *step*) was conducted in a *generalized linear model* (binomial logistic regression) for the binary response variable emotion recognition. The dispersion parameter of the quasi-binomial model and the nominal variable participants accounted for dependencies caused by repeated measurements within the participants.

In the global models, *stimulus types*, *acoustic parameters* and *emotions* were included as predictor variables. Participants, emotions, and stimulus types were fitted as nominal variables and acoustic parameters as quantitative variables. The order of the acoustic parameters in the models was determined by importance in a backward stepwise variable selection, that is in descending order starting with the acoustic parameter that explained most of the deviance. Conditional models were fitted for each stimulus type to account for interactions between stimulus types, emotion categories and acoustic parameters, since differences between fitted parameters of the models can be interpreted in terms of interactions. The relation between confidence ratings and emotion recognition was analyzed by a linear model with the response variable confidence ratings and the predictor variables

stimulus types and emotion recognition. Chi-square tests of the deviance analysis were used to analyze effects of predictor variables. In the quasi-binomial logistic regression, odds ratios were used to compare emotion categories as well as stimulus types. Confidence ratings of the linear model were compared by calculating the differences of the means. Tukey's method of multiple pairwise comparisons was used to compute simultaneous 95% confidence intervals for both, odds ratio and mean differences.

For the descriptive analysis of the data the following calculations were carried out: *relative frequencies*, *confusion matrices*, *classification errors* by random forest and listeners' judgements of emotion categories, *confidence intervals* by *binomial test* and Wagner's (1993) *unbiased hit rate* (H_u), which is the rate of correctly identified stimuli multiplied by the rate of correct judgments of the stimuli. The data was analyzed using the R language and environment for statistical computing and graphics version 3.4.3 (R Core Team, 2017) and the integrated environment *R-Studio* version 1.0.153 (used packages: *pwr*; *ipred*; *MASS*; *glm*; *step*; *multcomp*; *mvtnorm*; *lda*; *ggplot2*).

3.3 Results

3.3.1 Emotion category membership as predicted by acoustic parameters (LDA & RF)

The results obtained from LDA showed that the vast majority of variance was accounted by the first linear discriminant function [Anna (73.96%); pseudo-words (44.33%); semantic positive nouns (43.43%); semantic negative nouns (37.80%); semantic neutral nouns (51.01%); affect bursts (68.21%); pseudo-sentences (57.45%); lexical sentences (54.62%); neutral sentences (65.09%); across all stimuli in Group Words (47.06%); across all stimuli in Group Sentences (61.42%)]. This first discriminant function strongly correlated with *mean F0* [for Anna stimuli (r = -.708), pseudo-sentences (r = -.784), lexical sentences (r = -.755), neutral sentences (r = -.758) and across all stimuli in Group Sentences (r = -.799)], *duration* [for pseudo-words (r = .652), semantic negative nouns (r = -.782), semantic neutral nouns (r = -.699) and across all stimuli in Group Words (r = -.734)], *mean HNR* [for semantic positive nouns (r = .596)] and *amplitude* [for affect bursts (r = -.834)]. The second linear discriminant function accounted for 14.86% of the variance in Anna stimuli, 30.44% in pseudo-words, 27.80% in semantic positive nouns, 32.50% in semantic negative nouns, 22.47% in semantic neutral nouns, 15.41% in affect bursts, 18.84% in pseudo-sentences, 22.37% in lexical sentences, 16.79% in neutral sentences, 24.18% across all stimuli in Group Words and 14.99% across all stimuli in Group Sentences. This function correlated most strongly with *standard deviation HNR* [for Anna stimuli (r = -.720) and across all stimuli in Group Sentences (r = .519)], *duration* [for pseudo-words (r = -.696), affect bursts (r = -.790) and across all stimuli in Group Words (r = -.440)], *mean F0* [for semantic positive nouns (r = .663)], *mean HNR* [for semantic negative nouns (r = -.720)], *jitter* [for semantic neutral nouns (r= .667)], *standard deviation F0* [for pseudo-sentences (r = .533)], minimum F0 [for lexical sentences (r = -.561)] and *amplitude* [for neutral sentences (r = .585)]. ***Figure 2*** illustrates how the scores of the linear discriminant function

1 and linear discriminant function 2 separate the emotional categories for each stimulus type.[2] Comparisons between RF and LDA revealed that the error rates were overall smaller

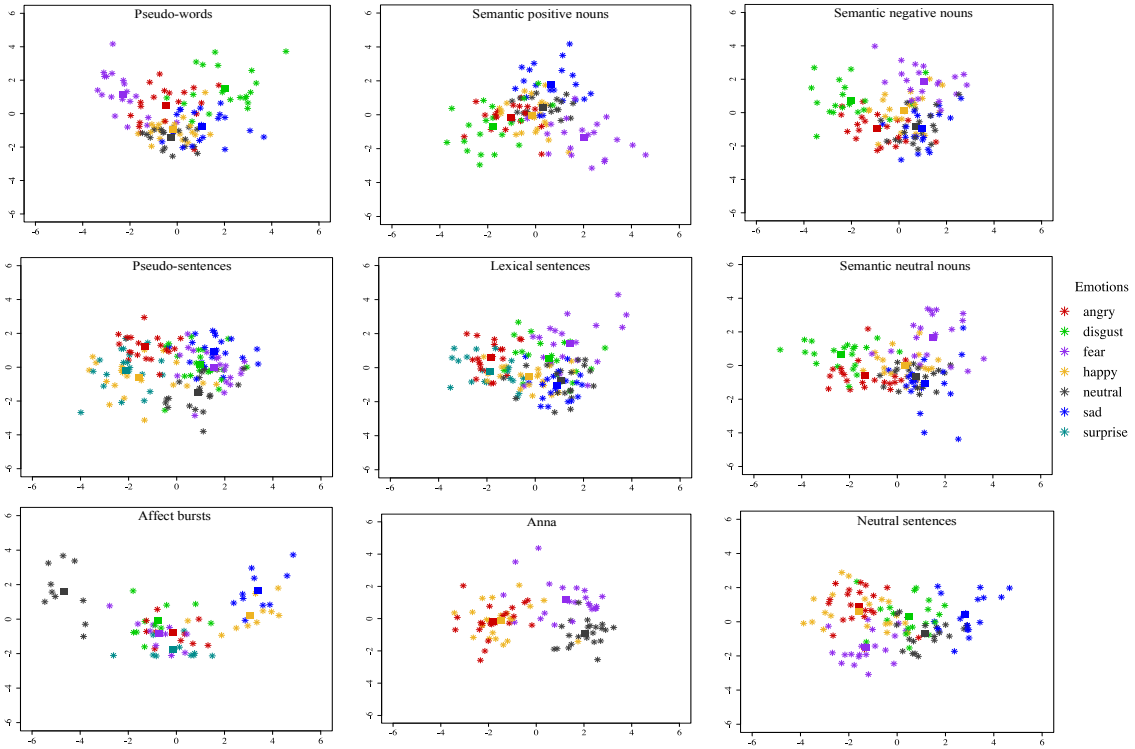


Figure 2 | Linear discriminant analysis for each stimulus type. On the ***x-axis*** is the *linear discriminant function 1* displayed, while on the ***y-axis*** the *linear discriminant function 2*. The squares represent the group means on each of the discriminant functions. Each stimulus is plotted according to its scores for the discriminant function 1 and function 2.

by RF than LDA when predicting emotion category membership across all 1038 stimuli. Specifically, error rates were reduced by 23.11% across all stimuli in *Group Words* and by 11.82% across all stimuli in *Group Sentences*. ***Table 1*** displays the error rates for both classification methods and the differences between RF and LDA error rates relative to the error rates of LDA.

Table 1 | Linear discriminant analysis (LDA) and random forest (RF) 10-fold cross-validation classification error rates for predicting vocal stimuli emotional category membership

Stimulus types	**Error rates**		
	LDA	***RF***	***Δ (%)****
Anna	0.3903	0.3778	3.20%
Pseudo-words	0.4597	0.4225	8.09%
Semantic positive nouns	0.4112	0.3428	16.63%
Semantic negative nouns	0.4614	0.4019	12.90%
Semantic neutral nouns	0.4402	0.4438	-0.82%
Group Words [across all stimuli (N = 568)]	0.4257	0.3754	11.82%
Affect bursts	0.4347	0.4701	-8.14%
Pseudo-sentences	0.5211	0.4259	18.27%
Lexical sentences	0.5866	0.4448	24.17%
Neutral sentences	0.4777	0.4255	10.93%
Group Sentences [across all stimuli (N = 470)]	0.5981	0.4813	19.53%
Overall [across all stimuli types (N = 1038)]	0.5802	0.4461	23.11%

Note: *The relative difference between RF error rates and LDA error rates were calculated as follows: **(1- (RF error rates / LDA error rates))*100**. As it can be observed, the error rates were smaller by RF than LDA, except for semantic neutral nouns and affect bursts. The accuracy rates for both classification methods can be obtained as follows: **(1 – error rates)*100**

[2] In *supplementary material* are presented (1) the complete tables for the correlations between acoustic parameters and all linear discriminant functions as well as the accounted variance for each function [*Table S1* (A_I to A_{XII})] and (2) the figures for the two linear functions that explained most of the variance within our stimuli datasets across all stimuli types in Group Words and Group Sentences [see *Fig. S1*]. In ***Appendix (A): Study 1*** are displayed these two functions by speakers' gender (see *Fig. A1* to *A3*).

3.3.2 Error classification patterns of emotions for listeners' judgments and RF algorithm

Table 2 displays the proportion of correctly identified emotions, Wagner's H_u, the comparisons in classification errors between listeners' judgments of emotion categories and RF algorithm, as well as, the $_{95\%}$CI of the exact binomial test.[3] As it can be observed, listeners

Table 2 | *Confusion matrices*, *unbiased hit rates* (H_u), comparisons of *classification errors* by random forest (*RF*) and listeners' judgments of emotion categories and *95%CI* of the exact *binomial test* for listeners classification errors

Stimulus type	Emotions portrayed	Emotion judgements									Classification errors	
		Angry	*Fear*	*Happy*	*Neutral*	*Sad*	*Disgust*	*Surprise*	*Total*	H_u	*Listeners* [$CI_{95\%}$]	*RF*
Anna	*Angry*	**2706**	134	24	93	23	89	121	3190	.688	15.17% [13.94%; 16.46%]	45.45%
	Fear	45	**1763**	41	231	637	42	431	3190	.429	44.73% [43.00%; 46.48%]	45.45%
	Happy	490	327	**919**	127	192	162	973	3190	.257	71.19% [69.59%; 72.76%]	50.00%
	Neutral	94	46	45	**2577**	373	40	14	3189*	.688	19.19% [17.84%; 20.60%]	18.18%
	Sad	—	—	—	—	—	—	—	—	—	—	—
	Disgust	—	—	—	—	—	—	—	—	—	—	—
	Surprise	—	—	—	—	—	—	—	—	—	—	—
	Total	3335	2270	1029	3028	1225	333	1539	12759	—	37.57% [36.73%; 38.42%]	37.78%
Pseudo-words	*Angry*	**2618**	47	50	21	14	65	85	2900	.693	9.72% [8.70%; 10.86%]	52.63%
	Fear	212	**1711**	93	104	530	37	408	2900	.503	41.00% [39.20%; 42.82%]	26.32%
	Happy	17	10	**1628**	699	13	16	477	2900	.384	43.86% [42.05%; 45.69%]	70.00%
	Neutral	57	22	150	**2155**	13	38	101	2899*	.410	25.66% [24.08%; 27.30%]	40.00%
	Sad	420	116	270	781	**1519**	37	89	2899*	.345	47.60% [45.77%; 49.44%]	35.00%
	Disgust	87	100	190	145	219	**1760**	273	2899*	.547	39.29% [37.51%; 41.09%]	25.00%
	Surprise	—	—	—	—	—	—	—	—	—	—	—
	Total	3411	2006	2381	3905	2308	1953	1433	17397	—	34.52% [33.82%; 35.24%]	42.25%
Semantic negative nouns	*Angry*	**2449**	46	24	23	11	273	74	2900	.639	15.55% [14.25%; 16.92%]	35.00%
	Fear	19	**2181**	73	15	350	29	232	2899*	.604	24.77% [23.20%; 26.38%]	25.00%
	Happy	33	178	**1862**	237	40	46	504	2900	.503	35.79% [34.05%; 37.57%]	80.00%
	Neutral	261	18	104	**2351**	29	60	77	2900	.530	18.93% [17.52%; 20.41%]	50.00%
	Sad	33	152	65	823	**1747**	24	55	2899*	.460	39.74% [37.95%; 41.55%]	10.00%
	Disgust	441	141	250	148	110	**1605**	205	2900	.436	44.66% [42.83%; 46.49%]	30.00%
	Surprise	—	—	—	—	—	—	—	—	—	—	—
	Total	3236	2716	2378	3597	2287	2037	1147	17398	—	29.91% [29.23; 30.59%]	40.19%
Semantic neutral nouns	*Angry*	**2565**	19	15	51	6	200	44	2900	.709	11.55% [10.41%; 12.77%]	65.00%
	Fear	17	**2137**	127	41	262	24	291	2899*	.650	26.28% [24.69%; 27.93%]	30.00%
	Happy	19	49	**1954**	333	18	16	511	2900	.536	32.62% [30.92%; 34.36%]	65.00%
	Neutral	204	5	62	**2555**	13	25	36	2900	.513	11.90% [10.74%; 13.13%]	75.00%
	Sad	18	88	97	1209	**1423**	15	49	2899*	.381	50.91% [49.08%; 52.75%]	30.00%
	Disgust	377	123	200	200	112	**1583**	305	2900	.464	45.41% [43.59%; 47.25%]	30.00%
	Surprise	—	—	—	—	—	—	—	—	—	—	—
	Total	3200	2421	2455	4389	1834	1863	1236	17398	—	29.78% [29.10%; 30.47%]	44.38%
Semantic positive nouns	*Angry*	**2390**	28	32	42	6	311	91	2900	.643	17.59% [16.22%; 19.02%]	35.00%
	Fear	14	**2138**	114	32	293	20	289	2900	.673	26.28% [24.68%; 27.92%]	35.00%
	Happy	75	18	**1884**	344	34	22	523	2900	.460	35.03% [33.30%; 36.80%]	35.00%
	Neutral	243	10	134	**2332**	30	34	117	2900	.503	19.59% [18.16%; 21.08%]	60.00%
	Sad	26	82	207	801	**1705**	19	59	2899*	.460	41.19% [39.39%; 43.00%]	15.00%
	Disgust	316	65	292	174	113	**1678**	262	2900	.466	42.14% [40.33%; 43.96%]	35.00%
	Surprise	—	—	—	—	—	—	—	—	—	—	—
	Total	3064	2341	2663	3725	2181	2084	1341	17399	—	30.30% [29.62%; 30.99%]	34.28%
Affect bursts	*Angry*	**940**	114	63	52	7	16	258	1450	.508	35.17% [32.71%; 37.69%]	90.00%
	Fear	96	**993**	9	38	10	52	252	1450	.455	31.52% [29.13%; 33.98%]	50.00%
	Happy	9	1	**1403**	7	10	0	20	1450	.869	3.24% [2.39%; 4.29%]	60.00%
	Neutral	20	5	11	**1384**	3	7	20	1450	.826	4.55% [3.54%; 5.75%]	10.00%
	Sad	1	11	44	1	**1387**	2	3	1449*	.925	4.28% [3.30%; 5.45%]	40.00%
	Disgust	76	9	10	66	12	**1204**	72	1449*	.719	16.91% [15.01%; 18.94%]	50.00%
	Surprise	58	361	22	51	6	110	**842**	1450	.333	41.93% [39.38%; 44.52%]	50.00%
	Total	1200	1494	1562	1599	1435	1391	1467	10148	—	19.66% [18.89%; 20.45%]	47.01%
Pseudo-sentences	*Angry*	**2420**	12	161	36	5	16	250	2900	.717	16.55% [15.22%; 17.96%]	30.00%
	Fear	53	**1642**	38	268	628	161	108	2898*	.393	43.34% [41.53%; 45.17%]	45.00%
	Happy	128	22	**1744**	109	5	12	880	2900	.275	39.86% [38.07%; 41.67%]	55.00%
	Neutral	7	9	92	**2660**	23	7	102	2900	.670	8.28% [7.30%; 9.34%]	30.00%
	Sad	12	359	45	98	**2260**	90	34	2898*	.533	22.02% [20.52%; 23.57%]	35.00%
	Disgust	91	292	162	453	376	**1371**	155	2900	.388	52.72% [50.89%; 54.55%]	55.00%
	Surprise	106	30	1578	18	8	12	**1148**	2900	.170	60.41% [58.61%; 62.20%]	55.00%
	Total	2817	2366	3820	3642	3305	1669	2677	20296	—	34.74% [34.09%; 35.40%]	42.59%
Lexical sentences	*Angry*	**2786**	3	24	12	3	14	58	2900	.835	3.93% [3.25%; 4.70%]	20.00%
	Fear	15	**2254**	28	85	404	12	102	2900	.731	22.28% [20.77%; 23.84%]	40.00%
	Happy	16	1	**2180**	365	3	5	330	2900	.397	24.83% [23.26%; 26.44%]	90.00%
	Neutral	30	21	40	**2658**	91	17	41	2899*	.684	8.31% [7.33%; 9.38%]	65.00%
	Sad	13	94	24	115	**2616**	18	19	2899*	.752	9.76% [8.71%; 10.90%]	55.00%
	Disgust	247	20	148	298	21	**1957**	208	2899*	.652	32.49% [30.79%; 34.23%]	35.00%
	Surprise	99	5	1686	28	2	3	**1077**	2900	.218	62.86% [61.07%; 64.62%]	40.00%
	Total	3206	2398	4130	3561	3140	2026	1835	20297	—	23.50% [22.91%; 24.09%]	44.48%
Neutral sentences	*Angry*	**2821**	3	34	23	0	5	14	2900	.851	2.72% [2.16%; 3.38%]	60.00%
	Fear	112	**1652**	250	155	25	30	676	2900	.425	43.03% [41.22%; 44.86%]	35.00%
	Happy	80	29	**2204**	75	24	16	472	2900	.666	24.00% [22.46%; 25.60%]	75.00%
	Neutral	141	22	11	**2498**	191	18	18	2899*	.647	13.83% [12.60%; 15.14%]	15.00%
	Sad	8	368	1	270	**2242**	5	6	2900	.605	22.69% [21.18%; 24.26%]	30.00%
	Disgust	61	141	14	306	385	**1958**	35	2900	.651	32.48% [30.78%; 34.22%]	65.00%
	Surprise	—	—	—	—	—	—	—	—	—	—	—
	Total	3223	2215	2514	3327	2867	2032	1221	17399	—	23.13% [22.50%; 23.76%]	42.55%
Group Words	*Angry*	**12728**	274	145	230	60	938	415	14790	.674	13.94% [13.39%; 14.51%]	41.58%
	Fear	112	**9930**	448	423	2072	152	1651	14788*	.567	32.85% [32.09%; 33.61%]	29.70%
	Happy	674	582	**8247**	1740	297	262	2988	14790	.422	44.24% [43.44%; 45.04%]	56.86%
	Neutral	1222	101	495	**11970**	458	197	345	14788*	.520	19.06% [18.43; 19.70%]	35.29%
	Sad	164	438	639	3614	**6394**	95	252	11596*	.358	44.86% [43.95%; 45.77%]	26.25%
	Disgust	1346	429	932	667	554	**6626**	1045	11599*	.458	42.87% [41.97%; 43.78%]	38.75%
	Surprise	—	—	—	—	—	—	—	—	—	—	—
	Total	16246	11754	10906	18644	9835	8270	6696	82351	—	32.13% [31.81%; 32.45%]	37.54%
Group Sentences	*Angry*	**8967**	132	282	123	15	51	580	10150	.758	11.66% [11.04%; 12.30%]	37.14%
	Fear	276	**6541**	325	546	1067	255	1138	10148*	.497	35.54% [34.61%; 36.48%]	48.57%
	Happy	233	53	**7531**	556	42	33	1702	10150	.465	25.80% [24.95%; 26.67%]	51.43%
	Neutral	198	57	154	**9200**	308	49	181	10147*	.688	9.33% [8.77%; 9.92%]	42.86%
	Sad	34	832	114	484	**8505**	115	62	10146*	.663	16.17% [15.46%; 16.90%]	42.86%
	Disgust	475	462	334	1123	794	**6490**	470	10148*	.583	36.05% [35.11%; 36.99%]	67.14%
	Surprise	263	396	3286	97	16	125	**3067**	7250	.180	57.70% [56.55%; 58.84%]	52.00%
	Total	10446	8473	12026	12129	10747	7118	7200	68139	—	26.18% [25.85%; 26.51%]	48.13%

Note: Frequencies of correctly judged portrayals are given on the main diagonal in boldface type. *If the number is less than the planned number of emotion judgments, this is due to few software errors in recording the emotion judgments. Classification errors of RF outside the 95%CI of listeners indicate a significant difference between the observed classification errors of listeners and RF.

most frequently misclassified *happy* for *surprise* (i.e., Anna, nouns, affect bursts, neutral

[3]Listeners recognition accuracy by stimulus type and emotions is displayed in *Tables S2(A)* and *S2(B)* in *supplementary material*.

sentences, pseudo-sentences) and *fear* for *sadness* (e.g., Anna, pseudo-words, semantic negative nouns, semantic positive nouns, lexical sentences, pseudo-sentences). In lexical- and pseudo-sentences, however, participants often mistook *surprise* for *happy*, whereas the *sad* tone of voice was frequently misclassified as *neutral* (e.g., pseudo-words, nouns, lexical sentences). Although generally well-recognized, utterances spoken in an *angry* tone of voice were mistaken for *surprise* (e.g., affect bursts, lexical- and pseudo-sentences, pseudo-words) and *disgust* (e.g., nouns), while for *neutral* and *disgusted* prosody no clear error pattern emerged [e.g., some utterances spoken in a *neutral* tone of voice were either mistaken for *angry* (Anna, nouns) or *sad* (lexical and neutral sentences), while utterances spoken in a *disgusted* tone of voice were misclassified as *angry* (nouns, affect bursts) or *neutral* (lexical-, pseudo-sentences)]. Comparing the proportion of classification errors between listeners' judgments of emotions and RF, one could observe that globally (i.e., across all emotions), humans were significantly better at predicting emotion category membership relative to RF, except for Anna stimuli, where no significant difference was observed. Looking at specific emotion categories, results indicated that in some stimulus sets the RF algorithm significantly outperformed listeners when classifying *disgust* and *sad* (i.e., pseudo-words and nouns), *fear* (i.e., pseudo-words and neutral sentences), *happy* (i.e., Anna stimuli) and *surprise* (i.e., pseudo-sentences and lexical sentences).

3.3.3 Emotion recognition and confidence ratings by stimulus type and emotion

The quasi-binomial and linear models revealed that stimuli types ($p < 0.001$), acoustic parameters (most of *p-values* < 0.001) and emotions ($p < 0.001$) significantly influenced listeners' performance accuracy of recognizing emotions and their confidence judgements. Moreover, results showed that listeners' confidence judgements were significantly affected by the correct identification of emotions (*p-values* < 0.001).[4]

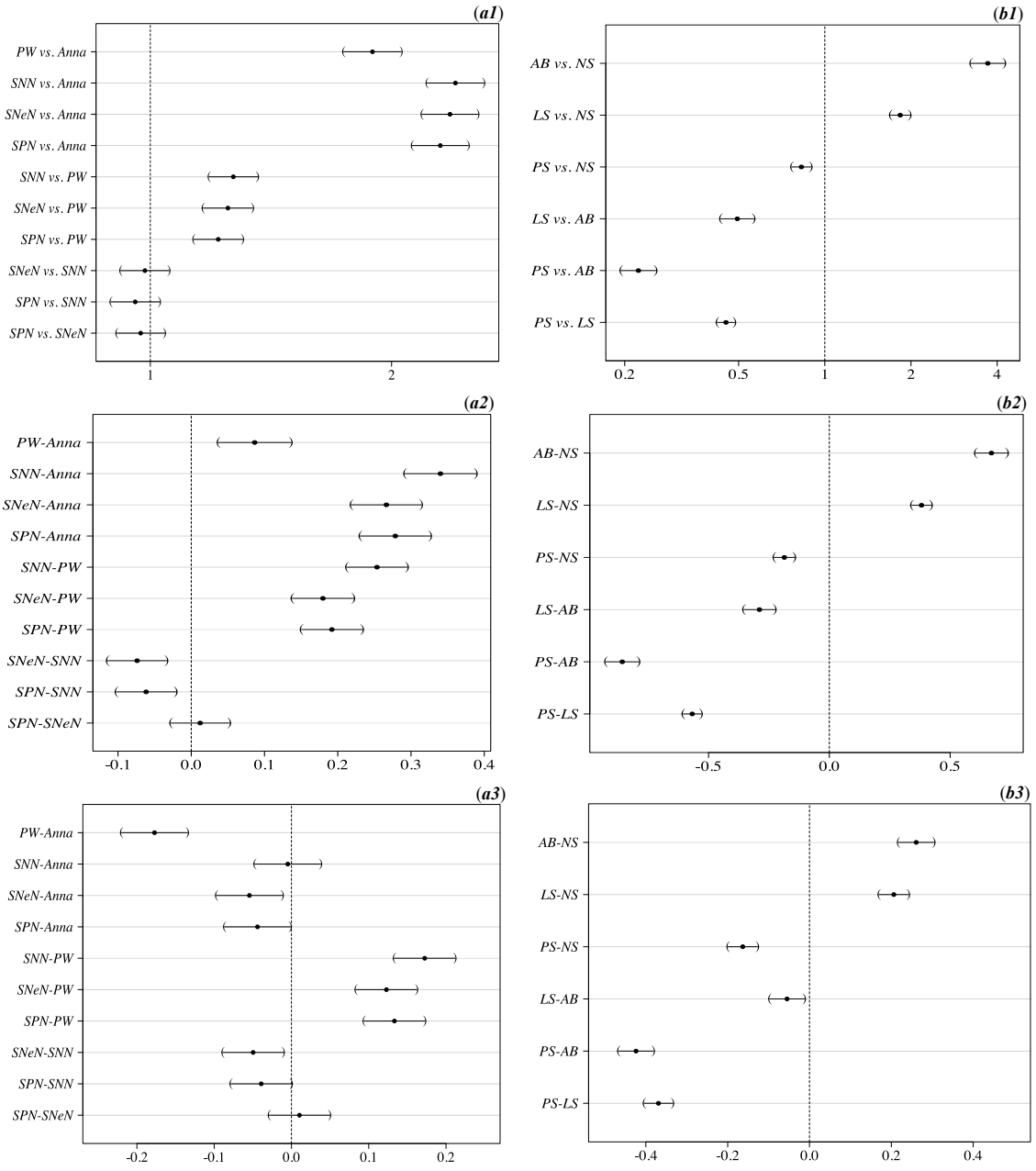
Odds ratio (*OR*) estimates indicated that in *Group Words* listeners were significantly more accurate at recognizing emotions in stimuli with a semantic connotation than in those spoken in a language devoid of meaning (semantic positive nouns vs. pseudo-words: $OR = 1.22$, $CI_{95\%}[1.13; 1.30]$; semantic negative nouns vs. pseudo-words: $OR = 1.27$, $CI_{95\%}[1.18; 1.36]$; semantic neutral nouns vs. pseudo-words: $OR = 1.25$, $CI_{95\%}[1.16; 1.34]$) or expressing a person's name (semantic positive nouns vs. Anna: $OR = 2.30$, $CI_{95\%}[2.12; 2.49]$; semantic negative nouns vs. Anna: $OR = 2.40$, $CI_{95\%}[2.21; 2.61]$; semantic neutral nouns vs. Anna: $OR = 2.37$, $CI_{95\%}[2.18; 2.56]$). The accuracy of performance when categorizing emotions was also significantly higher for *pseudo-words* than for *Anna* stimuli ($OR = 1.89$, $CI_{95\%}[1.74; 2.06]$). No significant differences in performance accuracy were found when comparing the stimuli with semantic content (semantic positive nouns vs. semantic neutral nouns: $OR = 0.97$, $CI_{95\%}[0.91; 1.04]$; semantic positive nouns vs. semantic negative nouns: $OR = 0.96, CI_{95\%}[0.89; 1.03]$; semantic neutral nouns vs. semantic negative

[4]To avoid a high degree of verbosity, the corresponding test statistics from the global- (i.e., across all stimuli types) and conditional models (i.e., for each stimulus type) are reported in *supplementary material* [see *Tables S2* (A, A_I, A_{II}) for Group Words, *S2* (B, B_I, B_{II}) for Group Sentences and *S2* (C, C_I, C_{II} to K, K_I, K_{II}) for each stimulus type].

nouns: OR = 0.98, $CI_{95\%}$[0.92; 1.06]). As shown by the multiple comparisons of the estimated means the pattern of the differences (Δ) in confidence judgments was similar to the pattern of recognition accuracies, however, listeners were less confident when identifying emotions in *semantic neutral-* (Δ = -0.07, $CI_{95\%}$[-0.11; -0.03]) and *semantic positive-* [(Δ = -0.06, $CI_{95\%}$[-0.10; -0.02]) than in *semantic negative nouns*. After adjusting for emotion recognition in the linear model, confidence ratings were significantly lower for *pseudo-words* (Δ = -0.18, $CI_{95\%}$[-0.22; -0.13]), *semantic neutral nouns* (Δ = -0.05, $CI_{95\%}$[-0.10; -0.01]) and *semantic positive nouns* (Δ = -0.04, $CI_{95\%}$[-0.09; -0.001]) than for *Anna* stimuli.

In *Group Sentences*, the odds of correctly identifying emotions, as well as, listeners' confidence judgments were significantly higher for *affect bursts* (OR = 3.71, $CI_{95\%}$[3.24; 4.24]; Δ = 0.67, $CI_{95\%}$[0.61; 0.74]) and *lexical sentences* (OR = 1.83, $CI_{95\%}$[1.70; 1.98]; Δ = 0.38, $CI_{95\%}$[0.34; 0.42]) than for *neutral sentences* and, lower for *lexical sentences* when compared to *affect bursts* (OR = 0.49, $CI_{95\%}$[0.43; 0.56]; Δ = -0.29, $CI_{95\%}$[-0.35; -0.23]). Recognition accuracy and confidence ratings were significantly lower for *pseudo-sentences* than for *affect bursts* (OR = 0.22, $CI_{95\%}$[0.19; 0.26]; Δ = -0.86, $CI_{95\%}$[-0.92; -0.79]), *lexical-* (OR = 0.45, $CI_{95\%}$[0.42; 0.48]; Δ = -0.57, $CI_{95\%}$[-0.60; -0.53]) and *neutral sentences* (OR = 0.83, $CI_{95\%}$[0.77; 0.89]; Δ = -0.19, $CI_{95\%}$[-0.23; -0.14]). This pattern remained similar even after adjusting for emotion recognition in the linear model (affect bursts – neutral sentences: Δ = 0.26, $CI_{95\%}$[0.22; 0.30]; lexical – neutral sentences: Δ = 0.21, $CI_{95\%}$[0.17; 0.24]; pseudo – neutral sentences: Δ = -0.16, $CI_{95\%}$[-0.20; -0.13]; lexical sentences – affect bursts: Δ = -0.05, $CI_{95\%}$[-0.10; -0.01]; pseudo-sentences – affect bursts: Δ = -0.42, $CI_{95\%}$[-0.47; -0.38]; pseudo – lexical sentences: Δ = -0.37, $CI_{95\%}$[-0.40; -0.33]). ***Figure 3*** illustrates the comparisons between stimuli types in both, *Group Words* and *Group Sentences*.

The comparison of performance accuracy for emotion categories showed that in both groups listeners were significantly less accurate and rated themselves as less confident when identifying emotional expressions spoken in a *disgusted*, *neutral*, *fearful*, *happy*, *sad* and *surprised* tone of voice than when spoken in an *angry* prosody (for emotion recognition with values ranging from $0.08 \leq OR \leq 0.83$; for confidence ratings with values ranging from $-0.90 \leq \Delta \leq -0.07$). Although in *Group Sentences*, listeners were significantly more accurate at categorizing utterances spoken in a *neutral-* than in an *angry* tone of voice (OR = 1.80, $CI_{95\%}$[1.55; 2.09]), the difference in confidence ratings was significantly lower for neutral than for angry expressions (Δ = -0.07, $CI_{95\%}$[-0.13; -0.00]). In addition, results showed that both, recognition rates and confidence ratings, were significantly higher when comparing *neutral* to other emotional prosodies (for emotion recognition with values ranging from $2.14 \leq OR \leq 7.65$; for confidence ratings with values ranging from $0.26 \leq \Delta \leq 0.83$). ***Figure 4*** illustrates the comparisons between emotion categories in *Group Words* and *Group Sentences* (see *Table S3* in *supplementary material* for the corresponding values). The pattern of results obtained for the comparisons between emotion categories for each type of stimulus are presented in *supplementary material* (see *Tables S4(A to I)* and the corresponding figures).



Stimuli types
PW = Pseudo-words
SNN = Semantic negative nouns
SNeN = Semantic neutral nouns
SPN = Semantic positive nouns
AB = Affect bursts
LS = Lexical sentences
NS = Neutral sentences
PS = Pseudo-sentences

Figure 3 | Comparisons between stimuli types in *Group Words* and *Group Sentences*
Emotion recognition odds ratio estimates for the comparisons between stimuli types are illustrated in panel (***a1***) and panel (***b1***). The linear contrasts for confidence ratings are illustrated in panel (***a2***) and (***b2***), while the confidence ratings after adjusting for emotion recognition are displayed in panel (***a3***) and (***b3***). Odds ratio of stimulus 1 (e.g., *AB*) vs. stimulus 2 (e.g., *NS*) less than 1 indicate that the recognition probability of stimulus 2 (e.g., *NS*) is higher than of stimulus 1 (e.g., *AB*), whereas values greater than 1 vice-versa. If the odds ratio of 1 is covered in the confidence interval, the difference in the recognition probabilities is not significant. Negative differences of confidence ratings of stimulus 1 (e.g., *AB*) vs. stimulus 2 (e.g., *NS*) indicate that the confidence ratings of stimulus 2 (e.g., *NS*) is higher than of stimulus 1 (e.g., *AB*), whereas positive differences vice-versa. If the difference of zero is covered in the 95%CI, the difference in the confidence ratings is not significant.

3.4 Discussion

The main goal of the present study was to investigate the influence of different types of vocal stimuli and their related acoustic parameters on emotional prosody recognition and retrospective confidence judgments. This was done by selecting a broad set of speech-embedded

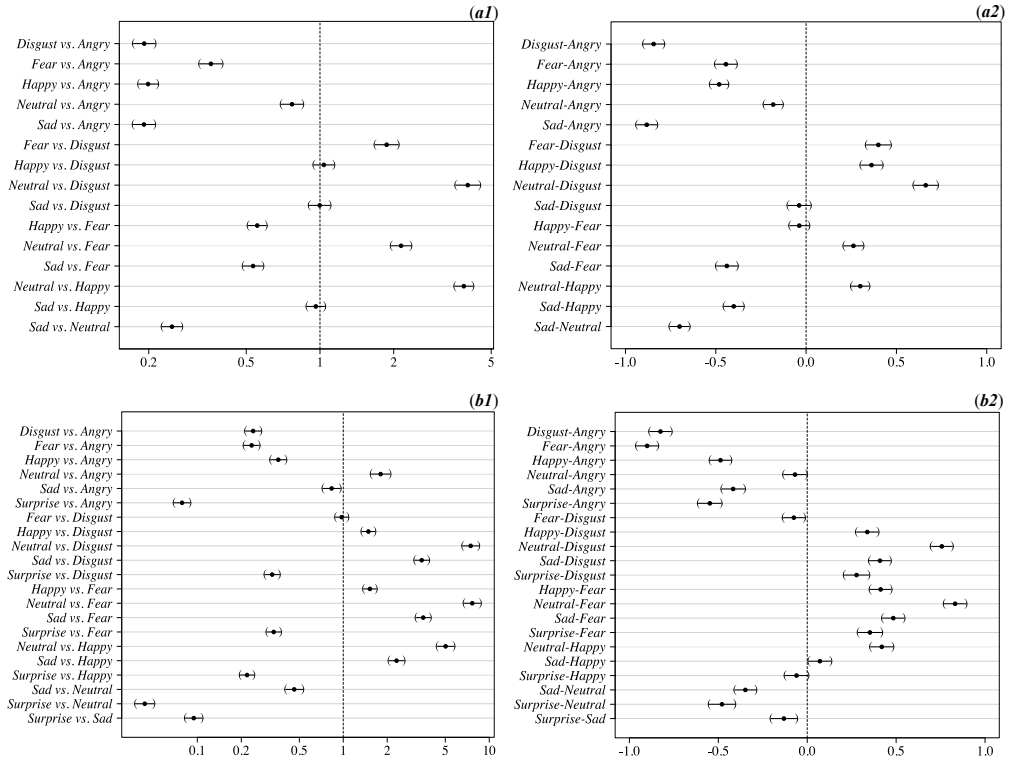


Figure 4 | Comparisons between emotion categories in *Group Words* (GW) and *Group Sentences* (GS)
The odds ratio estimates for the comparisons between emotion categories are illustrated in panel (***a1***) and panel (***b1***), while the linear contrasts for confidence ratings are illustrated in panel (***a2***) and (***b2***). Odds ratio of emotion 1 (e.g., *disgust*) vs. emotion 2 (e.g., *angry*) less than 1 indicate that the recognition probability of emotion 2 (e.g., *angry*) is higher than of emotion 1 (e.g., *disgust*), whereas values greater than 1 vice-versa. If the odds ratio of 1 is covered in the confidence interval, the difference in the recognition probabilities is not significant. Negative differences of confidence ratings of emotion 1 (e.g., *disgust*) vs. emotion 2 (e.g., *angry*) indicate that the confidence ratings of emotion 2 (e.g., *angry*) is higher than of emotion 1 (e.g., *disgust*), whereas positive differences vice-versa. If the difference of zero is covered in the 95%CI, the difference in the confidence ratings is not significant.

and non-speech embedded stimuli and implementing a standard acoustic parameter set based on previous findings highlighting the importance of paralingual emotional content in verbal communication. Specifically, we examined: (1) the extent to which listeners and classifiers use acoustic parameters as a perceptual cue for identifying portrayed emotions; (2) whether listeners' performance accuracy and confidence ratings are higher/lower for certain types of stimuli and, for specific emotion categories; (3) if correct recognition of emotions elicits confident judgments.

3.4.1 Performance accuracy grouped by classification algorithms & listeners

Overall, the findings provide additional support for the considerable body of work evidencing that different emotional states are signaled and communicated by specific acoustic characteristics (e.g., Johnstone & Scherer, 2000; Juslin & Laukka, 2003). Support, that there are specific vocal expression patterns for different emotions comes from two sets of findings. First, the linear discriminant algorithm was able to correctly classify the emotion category of 41.98% for all stimuli types unveiling specific constellations of predictors for each emotion and stimulus set with cross-validation estimates of accuracy ranging from 41.34% for lexical sentences to 60.97% for Anna stimuli. The results of this analysis compare well with previous work reporting accuracy rates for their stimulus materials between

40% and 57% (e.g., Banse & Scherer, 1996; Castro & Lima, 2010; K. Hammerschmidt & Jürgens, 2007; Sauter et al., 2010). By implementing RF as an additional classification method, using the same acoustic predictors, we observed that RF achieved a cross-validation classification accuracy across all stimuli, which was 31.94% relative higher than the accuracy of the LDA model. This result is in line with the findings reported by Noroozi et al. (2017) who investigated whether RF shows better predictive performance than deep neural networks (DNN) or more established techniques such as LDA, based on a set of 14 acoustic parameters extracted from *Surrey Audio-Visual Expressed Emotion database*. Their results showed that on average RF recognition rate was 26.25% higher relative to LDA and 11.02% higher compared to DNN. Although the comparison between different classification methods highlights that RF outperforms techniques such as LDA (Noroozi et al., 2017), we would like to note that results from both classification analyses demonstrated that acoustic measurements alone provide sufficient information to discriminate successfully between stimuli from different emotional categories. Using RF as a benchmark of listeners' performance accuracy, we observed that despite the fact that RF error patterns were significantly lower for certain emotions than those of listeners, overall (i.e., across all emotions) the automatic classification of emotions was considerably less successful than listeners' classification performance. Listeners' superior performance may be due to the fact that they could draw on a greater number of emotional markers (e.g., intonation patterns, emblems of distinct emotions [i.e., laughter, sighs], valence, arousal) inaccessible to the statistical algorithm, whose emotion category membership predictions were based on a fully automatic set of acoustic parameters extracted from relatively small learning datasets. Although classification algorithms do not seem to replicate the inference processes of human decoders, they appear to have lower error rates than listeners' when classifying certain emotions (e.g., disgust) solely based on their acoustic profiles. This has been shown by previous work (Banse & Scherer, 1996) as well as our current dataset. Second, the quasi-binomial logistic regression models demonstrated that the acoustic predictors accounted for a significant part of the deviance in recognition rates across all stimuli in the group listening to words (29.09%) and sentences (23.25%), as well as, for each type of stimulus with values ranging from 20.56% for neutral sentences to 67.27% for affect bursts.

Combined, these two sets of findings provide evidence for our first aim, showing that acoustic attributes of emotive speech explain a significant amount of variance in recognition rates and that the stimuli contained detectable acoustic contrasts which might have helped listeners to differentiate the portrayed emotion categories.

3.4.2 Emotion recognition and confidence ratings by stimulus type and emotion categories

Results from both logistic and linear models, showed that most of the acoustic predictors and vocal stimuli types had a significant influence on listeners' recognition of emotions and confidence judgements. Moreover, results showed that correct judgments of emotions elicited confident interpretations. These findings will be discussed in detail in the following.

The pattern of results within our study clearly indicated that listeners were signif-

icantly more accurate and confident at judging emotions from non-speech sounds (i.e., affect bursts) than speech-embedded stimuli (i.e., sentences/pseudo-sentences). This finding adds to previous research which demonstrated that affect bursts are decoded more accurately than speech-embedded prosody (Hawk et al., 2009). Further evidence comes from neurophysiological studies, showing that non-speech sounds facilitate early stages of perceptual processing in the form of decreased *N1* amplitudes and enhanced *P2* amplitudes (Liu et al., 2012; Pell et al., 2015). In other words, affect bursts, as evolutionary primitive signals, evoke a more rapid capture of attention than speech-embedded stimuli thought to involve more effortful cognitive processes and acoustic analysis (Pell et al., 2015). Moreover, as shown by our data, accurate decoding of affect bursts led to greater confidence judgments. In light of the above-mentioned studies, one could argue that these types of stimuli carry more ecologically relevant information and, thus are given precedence by the neurocognitive system, allowing individuals to be more accurate and confident in their judgments of desirable/undesirable events in their environment. In addition, our results demonstrated that for stimuli with lexico-semantic content (i.e., nouns, lexical and neutral sentences) the accuracy of performance and confidence ratings is significantly higher than for stimuli devoid of meaning (i.e., pseudo-words; pseudo-sentences). Support for these findings comes from validation studies showing greater accuracy and higher ratings for semantic- compared to pseudo-utterances (e.g., Castro & Lima, 2010). Similar to previous research, we also found that lexical sentences, which were based on congruent combinations of semantics and prosody, yield not only greater accuracy (e.g., Ben-David et al., 2016) but also higher confidence ratings compared to neutral sentences. When comparing stimuli with a semantic positive, negative and neutral content, no significant differences in recognition accuracy were observed, yet, listeners felt significantly more confident at detecting emotions in semantic negative- than semantic positive or neutral nouns. One explanation that has been put forth as to why such a negativity bias occurs in social judgments is that people may generally consider negative information to be more diagnostic than positive information in forming an overall impression (e.g., Hamilton & Huffman, 1971). This is supported by studies showing that people consider negative information to be more important to impression formation and, when it is available to them, they are subsequently more confident (e.g., Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Hamilton & Zanna, 1972). Another possible explanation is that negative stimuli exert a stronger influence on people's evaluations because they are more complex than positive ones and, thus, require greater attention and cognitive processing (e.g., Abele, 1985; Ito, Larsen, Smith, & Cacioppo, 1998; Peeters & Czapinski, 1990). As shown by our data, correct recognition determines an increase of confidence judgments, however, when comparing Anna stimuli to other stimuli types, we observed that, when correctly categorizing the emotions for this stimulus type, listeners' felt more confident regarding the correctness of their answer. As their decision was based on hearing the same item (i.e., the name Anna) repeatedly, one could argue that this might have led to a familiarity effect with the stimulus. Studies in the metacognitive literature have shown item familiarity leads to higher confidence ratings, because participants rest on the belief that more knowledge about the item means they

are more accurate, although this has been shown to be an irrelevant factor (e.g., Koriat, 2008; Metcalfe, Schwartz, & Joaquim, 1993).

Another finding within the current study was that, for a vast majority of the stimuli, listeners' performance accuracy and confidence ratings were significantly higher when spoken in an angry and neutral tone of voice than in any other emotional prosody.In contrast, recognition accuracy and confidence ratings were lowest for disgust, excepting affect bursts. One interesting pattern that emerged from our data was that listeners felt more confident

Table 3 | Schematic overview for all emotion comparisons by stimulus type and across all types of stimuli

		Anna		PW		SPN		SNN		SNeN		AB		PS		LS		NS		GW		GS	
		ER	*CR*	*ER*	*CR*	*ER*	*CR*	*ER*	*CR*	*ER*	*CR*	*ER*	*CR*	*ER*	*CR*	*ER*	*CR*	*ER*	*CR*	*ER*	*CR*	*ER*	*CR*
D	A																						
F	A							ns.	ns.			ns.	ns.										
H	A																						
N	A	ns.				ns.	ns.			ns.					ns.	ns.							
S	A															ns.							
Su	A											ns.	ns.										
F	D			ns.	ns.												ns.					ns.	
H	D									ns.			ns.					ns.		ns.			
N	D																						
S	D						ns.	ns.			ns.		ns.								ns.	ns.	
Su	D																						
H	F			ns.	ns.	ns.	ns.				ns.			ns.		ns.					ns.		
N	F						ns.	ns.															
S	F																						
Su	F												ns.										
N	H						ns.		ns.			ns.							ns.					
S	H					ns.						ns.	ns.		ns.					ns.			
Su	H																						ns.
S	N											ns.	ns.			ns.							
Su	N																						
Su	S														ns.								

Note: *ER* = Emotion recognition; *CR* = Confidence ratings; *A* = Angry; *D* = Disgust; *F* = Fear; *H* = Happy; *N* = Neutral; *S* = Sad; *Su* = Surprise; *PW* = Pseudo-words; *SPN* = Semantic positive nouns; *SNN* = Semantic negative nouns; *SNeN* = semantic neutral nouns; *AB* = Affect bursts; *PS* = Pseudo-sentences; *LS* = Lexical sentences; *NS* = Neutral sentences; *GW* = Group Words; *GS* = Group Sentences. The color represents the emotion with the higher recognition rates and confidence ratings; *ns.* = not significant.

at categorizing surprise than disgust and fear (i.e., pseudo- and lexical sentences) or happy (i.e., lexical sentences), despite the fact that their performance accuracy was significantly lower for this emotion. Moreover, we found that for nouns, fear had higher accuracy rates and confidence ratings than sadness, yet, in the other types of stimuli the exact opposite pattern was observed. Happiness yielded higher accuracy scores and confidence ratings in comparison to disgust, while when compared to other emotions this largely depended on the type of stimulus (see, ***Table 3*** for a schematic overview for all emotion comparisons by stimulus type and across all types of stimuli).

Findings from emotion prosody literature provide conflicting evidence regarding the differences between emotional and neutral prosody. Past work has demonstrated that stimuli spoken in a neutral tone of voice are identified more accurately and have higher confidence ratings when compared to other emotional prosodies, regardless of semantics (Cornew, Carver, & Love, 2010; Schirmer & Kotz, 2003). Our findings and previous work on vocal emotion recognition converge not only towards a general advantage for recognizing neutral prosody, but also angry expressions (e.g., Chronaki et al., 2018; Paulmann & Uskul, 2014; Pell, Monetta, Paulmann, & Kotz, 2009; K. R. Scherer et al., 2001). This is also compatible with evolutionary theories arguing that humans (and other primates) are biologically prepared to respond rapidly to vocal cues associated with threat or anger (e.g., Öhman, 1993). It has been argued that non-speech sounds (e.g. growls of anger, the laughter of happiness or cries of sadness) convey emotions clearer and faster than words (Pell et al., 2015). However, our results showed that when identifying emotions from affect bursts, listeners were less accurate and felt less confident at detecting anger when compared to other emotional prosodies. Previous findings report similar accuracy patterns when decoding anger from affect bursts (Belin et al., 2008; Pell et al., 2015). However, it

remains unclear why this effect emerged. A possible explanation relates to the acoustics of acting anger sounds which might differ from natural ones. Anikin and Lima (2018), for instance, showed that authentic vocalizations (e.g., anger, fear) differ from actor portrayals in a number of acoustic characteristics by showing a higher pitch and lower harmonicity, as well as, a less variable spectral slope and amplitude. Thus, it is plausible that these acoustic characteristics of authenticity are hard-to-fake markers of a speaker's emotional state and thus signal a distinction between honest communication and a bluff (Anikin & Lima, 2018). The patterns obtained for disgust are also consistent with those of previous studies showing that, after surprise, it is the most difficult emotional display to recognize from speech-embedded stimuli (e.g., Paulmann & Uskul, 2014) but not from non-speech sounds, presumably, because this emotion is often expressed in affect bursts or short interjections (e.g., yuck) rather than in sentential context (Banse & Scherer, 1996; Johnstone & Scherer, 2000). In contrast, surprise yielded higher confidence scores than disgust or fear, which could be due to the fact that humans are more prone to notice and focus on surprising events and, therefore are more likely to attend to them (Wilson & Gilbert, 2008). Going further, one could speculate that similar to anger, surprise might also serve a functional and adaptive purpose as people might devote their energy to judging whether what is unfolding before them is a threat, a joke or a harmless event, thus, eliciting more confident evaluations. A similar argument may apply to our results regarding the comparisons between fear and sadness. Fear, as an expression that signals threat, might require less auditory input to be decoded accurately (i.e., shorter stimuli – in our case nouns), while identifying sadness from speech might activate additional social meanings that take more time to analyze and more careful post-message processing.

To summarize, this set of results extends previous findings from the facial domain (Kelly & Metcalfe, 2011) by showing that listeners who were better at recognizing vocal expressions of emotion were also more confident in their judgments. Although slight variations between emotion recognition accuracy and confidence ratings were observed for some stimuli types or emotion categories, overall our results demonstrate that the correct recognition of emotions elicits confident judgments. This suggests that individuals can predict and assess their performance for the recognition of emotional prosody.

3.4.3 Limitations & Future Research

Although the main groups of paralinguistic features and their acoustic parameters were covered (Eyben et al., 2016; Juslin & Laukka, 2003), not all relevant properties of vocal emotional expression have been considered. In a single contribution, however, this may posit difficulties due to space limitations. Nevertheless, future research would profit by implementing, for instance, spectral parameters (e.g., alpha ratio, Hammarberg index) or prosodic contours, as it has been argued they index physiological changes in voice and are sensitive to emotional expressions (e.g., Eyben et al., 2016; Mozziconacci, 2002). A related limitation is the fact that parameters were extracted from the whole utterance. Although this is a common approach (e.g., Castro & Lima, 2010; Paulmann et al., 2008; Pichora-Fuller et al., 2016), it has been suggested that it disregards the phonetic identity

of speech segments on emotional expression (Goudbeek & Scherer, 2010). There is also the proviso that gender of the speaker may have had an effect on the discrimination accuracy. However, the present study focused on the patterns of voice cues used to portray specific emotions, rather than on gender differences. By keeping in line with previous research in this area we extracted the acoustic parameters across genders (e.g., Juslin & Laukka, 2001; Paulmann et al., 2008; Sauter et al., 2010). This does not rule out that gender might have had an effect, however, it should be noted that for the majority of the stimulus types there were only two speakers (1 male, 1 female). Further work with a greater number of speakers would not only be able to establish the degree to which the acoustic factors in this study can be generalized, but would also help to explain the variation in these factors alongside speakers' gender characteristics. A further limitation regards the absence of some emotional categories within our stimuli datasets. In comparison to the classification algorithms which categorized emotions based on the existing number of emotion categories, listeners were supposed to choose from a fixed set of given alternatives. This might explain why for certain emotion categories, listeners had higher error rates than the RF algorithm. Finally, the sample in the current study was limited to a university-educated population and included predominantly young adults, which may limit the generalizability of the findings to the wider population.

3.4.4 Conclusion

This study provides the first systematic investigation of the influence acoustic parameters and stimulus types exert on vocal emotion recognition and metacognition. The findings within the present study are essential both empirically and conceptually. First, they replicate earlier research findings by establishing that humans can infer emotion from vocal expression alone, based on differential acoustic patterning. Second, our results add to previous findings by demonstrating that emotional expressions are more accurately recognized and confidently judged from non-speech sounds than from emotionally inflected speech. In addition, they answer previously unaddressed research questions (Sauter et al., 2010) by showing that this pattern is not constant across all emotional categories and that listeners do not rely on the same acoustic cues when decoding emotions from speech and non-speech embedded sounds. While the current findings demonstrate that correct recognition of emotions promotes confident interpretations, more research is needed to uncover the underlying mechanisms of how individuals use this metacognitive knowledge.

Funding

This research was funded by the *Deutsche Forschungsgemeinschaft* (DFG, German Research Foundation) – Project number 254142454/GRK 2070.

Acknowledgments

We would like to thank Silke Paulmann for generously providing us with her stimuli sets, Carlotta Dove, Isabel Nöthen and Christina Broering for help with data acquisition and all individuals who participated in the research presented here.

3.5 Supplementary Material

Processing of acoustic parameters

The processing of acoustic parameters was made based on the script packages developed by Owren (2008). The extraction of duration, peak time, amplitude and peak amplitude was made using "Quantify Amplitude and Duration" script. For the extraction of F0-related parameters, HNR parameters and shimmer, we used the "Quantify Source" script. The processing of stimuli was made using the objects 'no labels' mode and by applying the default settings from Owren's 2008 script packages. The default settings in the first script, set the minimum frequency at which the program would take an intensity measurement to 100 Hz, with a standard time step of 0.0 seconds. In the second script, the pitch floor is set to 75Hz and the pitch ceiling to 600Hz for all speakers. For jitter, the relative average perturbation is provided, while shimmer is determined by the average absolute base-10 logarithm of the difference between the amplitudes of consecutive periods multiplied by 20 (for the mathematical formulas on these two functions, see for example Teixeira & Gonçalves, 2014). The harmonicity parameters were set to match the pitch extraction (see Praat Manual, for details http://www.fon.hum.uva.nl/praat/manual/).

Random Forest (RF)

The random forest algorithm is built on several concepts: decision or classification trees, bootstrap resampling, majority voting, and random variable selection at each node of a tree. Classification trees are computed by a recursive algorithm. The algorithm starts with a decision rule which partitions the learning sample in two subgroups and is applied to each subgroup again until a stop criterium is fulfilled. The decision rule is defined by a cut-point in one of the parameters, which assigns all observations to two different subgroups, depending on the parameter's value. The algorithm chooses the split defined by the cut-point and parameter which maximizes a measure of information gain or diversity between both subgroups. A stop criterium of the classification tree algorithm is the size of the subgroup. If the algorithm stops for a subgroup, the subgroup becomes a leaf of the tree. New observations are classified with the classification tree by identifying the leaf of the new observation. The majority class of the observations of the learning sample in the identified leaf is the predicted class of the new observation. It is known that single classification trees are weak and unstable classifiers. To achieve a more stable classifier, bootstrap resampling is used to compute a sample of classification trees. The size of the sample M is a hyperparameter of the method. Often values between 200 and 1000 are used for M. A classification tree is computed for each bootstrap sample. The set or forest of classification trees defines the ensemble classifier, named by the acronym bagging (bootstrap aggregation). A new observation is classified by the M classification trees of the bagging ensemble. The majority class of the M classifications defines the predicted class by the bagging classifier. RF increases the instability of each single classification tree in the ensemble by introducing a random variable selection method at each node (recursion) of the classification tree algorithm. The random forest classification tree algorithm chooses

at each node of the tree the cut-point and parameter which maximizes a measure of information gain or diversity between both subgroups out of a random subset of available parameters. Bootstrap aggregation and majority vote are used to define the random forest ensemble classifier (see Breiman, 2001; James et al., 2013, for more details). The hyperparameters of RF were chosen by the bootstrap sample size of 500 and by randomly selecting 3 out of 13 acoustic parameters (the default values of the R package *randomForest*) to determine the optimal split at each node.

LDA: The results of all linear discriminant functions and the accounted variance for each function

Table S1 (A_I) | Correlations between acoustic parameters and linear discriminant functions for *Anna* stimuli ($N_{Stimuli}$ = 88)

	Duration	*PeakTime*	*Amp.(dB)*	*PeakAmp.*	*Min.F₀*	*Max.F₀*	*Mean.F₀*	*StDev.F₀*	*Jitter*	*Shimmer*	*Max.HNR*	*Mean HNR*	*StDev. HNR*	*Accounted variance*
LD1	-0.6550	-0.4105	0.5178	-0.3243	-0.3689	-0.6487	-0.7077	-0.6191	0.4347	0.0731	-0.2555	0.5794	-0.0686	73.96%
LD2	-0.2884	-0.5230	0.0302	0.0695	-0.4311	-0.2994	-0.3531	-0.1385	0.3239	0.1730	-0.3151	-0.4285	-0.7203	14.86%
LD3	-0.2517	0.0481	-0.4371	0.0072	0.3591	0.0672	0.2086	-0.1060	-0.1276	0.0576	-0.2166	-0.2475	-0.3345	11.18%

Table S1 (A_{II}) | Correlations between acoustic parameters and linear discriminant functions for *Pseudo-words* ($N_{Stimuli}$ = 120)

	Duration	*PeakTime*	*Amp.(dB)*	*PeakAmp.*	*Min.F₀*	*Max.F₀*	*Mean.F₀*	*StDev.F₀*	*Jitter*	*Shimmer*	*Max.HNR*	*Mean HNR*	*StDev. HNR*	*Accounted variance*
LD1	0.6516	0.2088	-0.2930	0.1594	-0.3676	0.0779	-0.2259	0.3013	0.1999	0.2833	0.0713	-0.4523	-0.0370	44.33%
LD2	-0.6959	-0.6371	-0.1768	-0.3871	-0.5873	-0.3734	-0.5824	0.0872	0.5088	0.4769	-0.1575	-0.4668	-0.3035	30.44%
LD3	0.1601	-0.1624	0.1693	-0.3448	-0.0091	0.3264	0.0009	0.4490	0.1900	0.4313	0.4699	0.3566	0.6204	21.81%
LD4	-0.0649	-0.1455	-0.3100	0.2262	0.0213	-0.4008	-0.5052	-0.5841	-0.0001	-0.3424	0.3013	0.3769	-0.1583	2.97%
LD5	0.0236	-0.0671	-0.2387	-0.3757	-0.2814	0.1135	0.0309	0.0563	-0.1906	-0.1428	0.3513	-0.0242	0.2062	0.45%

Table S1 (A_{III}) | Correlations between acoustic parameters and linear discriminant functions for *Semantic positive nouns* ($N_{Stimuli}$ = 120)

	Duration	*PeakTime*	*Amp.(dB)*	*PeakAmp.*	*Min.F₀*	*Max.F₀*	*Mean.F₀*	*StDev.F₀*	*Jitter*	*Shimmer*	*Max.HNR*	*Mean HNR*	*StDev. HNR*	*Accounted variance*
LD1	-0.5880	-0.5140	0.4371	-0.4764	0.4900	0.0003	0.2144	-0.1764	-0.0530	-0.1635	0.3862	0.5964	0.3354	43.43%
LD2	0.1597	0.3553	-0.0569	0.3261	0.6262	0.4414	0.6631	0.0075	-0.6117	-0.1554	0.1100	0.2606	-0.0219	27.80%
LD3	-0.4239	-0.4421	-0.3022	0.2163	-0.0504	-0.3838	0.0626	-0.2874	-0.0878	-0.0064	-0.0432	-0.0412	-0.1541	16.24%
LD4	-0.4934	-0.0106	0.2463	0.1016	0.0683	-0.2153	-0.2440	-0.2215	0.2368	0.5420	-0.3328	-0.4850	-0.4538	9.52%
LD5	0.1951	-0.0586	-0.6162	-0.2853	0.3609	0.0930	0.2685	-0.0097	0.3131	0.1877	0.0364	-0.0801	-0.1004	3.01%

Table S1 (A_{IV}) | Correlations between acoustic parameters and linear discriminant functions for *Semantic negative nouns* ($N_{Stimuli}$ = 120)

	Duration	*PeakTime*	*Amp.(dB)*	*PeakAmp.*	*Min.F₀*	*Max.F₀*	*Mean.F₀*	*StDev.F₀*	*Jitter*	*Shimmer*	*Max.HNR*	*Mean HNR*	*StDev. HNR*	*Accounted variance*
LD1	-0.7820	-0.3812	0.3580	-0.4450	0.1882	-0.0908	0.0926	-0.1394	-0.0887	-0.1123	0.4151	0.5951	0.3042	37.80%
LD2	-0.3474	-0.1614	0.2205	-0.2892	-0.4186	-0.3401	-0.4058	-0.0113	0.4711	0.3990	-0.2325	-0.7195	-0.4151	32.50%
LD3	0.2931	0.0506	0.1193	-0.3933	0.1040	0.4269	0.2834	0.4927	0.1921	0.4596	0.2591	0.0143	0.0290	23.98%
LD4	0.0243	0.4241	0.4494	0.0440	0.2619	-0.2196	-0.1234	-0.3211	-0.0529	0.1399	-0.1836	-0.0199	0.1601	4.94%
LD5	-0.1091	0.3389	-0.2069	-0.2894	0.5587	0.0836	0.4735	-0.1444	-0.1949	0.0133	0.0199	-0.2852	-0.4258	0.77%

Table S1 (A_V) | Correlations between acoustic parameters and linear discriminant functions for *Semantic neutral nouns* ($N_{Stimuli}$ = 120)

	Duration	*PeakTime*	*Amp.(dB)*	*PeakAmp.*	*Min.F₀*	*Max.F₀*	*Mean.F₀*	*StDev.F₀*	*Jitter*	*Shimmer*	*Max.HNR*	*Mean HNR*	*StDev. HNR*	*Accounted variance*
LD1	-0.6988	-0.6441	0.1950	-0.3956	0.1780	-0.2840	0.0326	-0.2611	0.0312	-0.1568	0.3224	0.6066	0.3753	51.01%
LD2	-0.3966	-0.3451	-0.0595	-0.2619	-0.6579	-0.1952	-0.5096	0.2279	0.6669	0.5987	0.0228	-0.6520	-0.3165	22.47%
LD3	0.3598	0.1220	0.1226	-0.2588	0.1393	0.1736	0.0982	0.1536	0.2564	0.3875	0.2628	0.2269	0.2799	17.31%
LD4	-0.2653	0.0920	0.3104	0.2565	0.2511	-0.0532	-0.0828	-0.2179	0.0144	0.2796	-0.0939	-0.1927	-0.0877	6.88%
LD5	0.0624	0.0817	-0.3346	0.1144	0.1838	-0.0969	0.2192	-0.0545	0.1728	0.0511	-0.4377	0.0735	-0.0411	2.34%

Table S1 (A_{VI}) | Correlations between acoustic parameters and linear discriminant functions for *Affect bursts* ($N_{Stimuli}$ = 70)

	Duration	*PeakTime*	*Amp.(dB)*	*PeakAmp.*	*Min.F₀*	*Max.F₀*	*Mean.F₀*	*StDev.F₀*	*Jitter*	*Shimmer*	*Max.HNR*	*Mean HNR*	*StDev. HNR*	*Accounted variance*
LD1	0.4977	0.4064	-0.8339	0.6568	0.0305	0.5406	0.4088	0.5719	0.5084	0.7182	0.3826	-0.5260	-0.1476	68.21%
LD2	-0.7900	-0.4511	0.1195	-0.3553	0.4308	-0.0031	0.4103	-0.0435	-0.2068	0.0510	-0.3888	-0.4061	0.0698	15.41%
LD3	0.0696	-0.0182	-0.2587	0.1858	-0.5911	-0.1834	-0.3875	0.2184	-0.0148	0.0907	-0.2950	-0.3951	-0.5560	8.31%
LD4	0.1661	-0.0403	0.1205	-0.2582	-0.0837	0.1931	0.1337	0.0806	-0.4611	-0.4945	0.0049	0.2333	0.6441	4.84%
LD5	0.0088	-0.1663	-0.1538	-0.0281	0.0017	-0.2499	-0.3898	-0.4222	0.2124	0.1332	-0.4542	-0.4561	-0.3228	2.32%
LD6	-0.1856	-0.0637	-0.0601	0.0177	-0.3114	0.0683	-0.3525	0.1975	0.1065	-0.2030	0.3413	0.0892	0.2451	0.9%

Table S1 (A_{VII}) | Correlations between acoustic parameters and linear discriminant functions for *Pseudo-sentences* ($N_{Stimuli}$ = 140)

	Duration	*PeakTime*	*Amp.(dB)*	*PeakAmp.*	*Min.F₀*	*Max.F₀*	*Mean.F₀*	*StDev.F₀*	*Jitter*	*Shimmer*	*Max.HNR*	*Mean HNR*	*StDev. HNR*	*Accounted variance*
LD1	0.3771	-0.0935	-0.1265	0.0558	-0.4494	-0.3408	-0.7839	-0.6346	0.4642	0.1988	-0.0966	0.0223	0.4643	57.45%
LD2	0.1297	-0.4157	-0.2376	0.1868	-0.1973	0.2129	-0.0873	0.5332	0.3590	0.1198	0.2794	-0.3154	0.0816	18.84%
LD3	-0.4599	-0.1373	0.2219	-0.1013	-0.4693	-0.1857	-0.1647	-0.0117	0.0714	-0.1961	-0.1996	0.1493	-0.1164	12.31%
LD4	-0.6942	-0.2113	0.2765	-0.1699	-0.0671	-0.1698	-0.1086	-0.0720	0.1412	0.5983	-0.0209	-0.4359	-0.3130	5.71%
LD5	0.1445	-0.4182	-0.0886	0.2827	-0.3038	-0.3042	-0.2701	-0.2797	0.1111	-0.0912	-0.3107	-0.4274	-0.3197	4.50%
LD6	0.2018	-0.0895	0.6162	-0.0777	0.0905	-0.1101	-0.0415	-0.0159	0.0344	-0.1563	-0.5095	0.0604	0.0731	1.19%

Table S1 (A_{VIII}) | Correlations between acoustic parameters and linear discriminant functions for *Lexical sentences* ($N_{Stimuli}$ = 140)

	Duration	*PeakTime*	*Amp.(dB)*	*PeakAmp.*	*Min.F₀*	*Max.F₀*	*Mean.F₀*	*StDev.F₀*	*Jitter*	*Shimmer*	*Max.HNR*	*Mean HNR*	*StDev. HNR*	*Accounted variance*
LD1	0.4448	-0.0538	-0.1066	-0.2495	-0.4117	-0.2307	-0.7554	-0.1422	0.5294	0.3298	-0.4353	-0.1898	0.2236	54.62%
LD2	-0.4782	-0.3829	0.3346	-0.0880	-0.5608	-0.1887	-0.3494	-0.1514	0.1789	-0.2302	-0.0985	0.0331	0.0491	22.37%
LD3	0.0127	0.2767	0.1037	0.3610	-0.1501	-0.5611	-0.2546	-0.6606	0.1261	-0.0430	-0.3907	-0.1173	-0.3063	14.99%
LD4	0.0809	0.4564	0.1510	0.3253	-0.1189	0.0613	0.0991	0.1869	0.2217	0.2106	0.0965	0.1215	0.4180	4.27%
LD5	-0.0359	-0.1000	-0.3284	-0.2138	0.2896	-0.1926	0.2997	-0.2107	-0.4460	-0.2428	-0.3589	0.5963	0.3700	2.16%
LD6	0.0052	-0.1090	-0.6472	0.3348	0.0077	-0.2307	-0.0753	-0.1331	-0.0609	0.1186	0.3954	0.0211	0.0686	1.60%

Table S1 (A_{IX}) | Correlations between acoustic parameters and linear discriminant functions for *Neutral sentences* ($N_{Stimuli}$ = 120)

	Duration	*PeakTime*	*Amp.(dB)*	*PeakAmp.*	*Min.F₀*	*Max.F₀*	*Mean.F₀*	*StDev.F₀*	*Jitter*	*Shimmer*	*Max.HNR*	*Mean HNR*	*StDev. HNR*	*Accounted variance*
LD1	0.3477	-0.0270	-0.0112	0.0505	-0.5353	0.0839	-0.7577	0.2029	0.4682	0.1322	-0.3693	0.1758	0.1648	65.09%
LD2	-0.3904	-0.4358	0.5850	0.2705	0.2088	-0.0268	-0.0753	-0.2574	0.0738	0.0741	-0.3325	-0.0078	-0.4617	16.79%
LD3	0.1694	0.2987	-0.3042	-0.1668	0.5974	0.3558	0.3997	0.2659	-0.1578	-0.1997	0.1440	0.3144	0.4113	10.12%
LD4	0.4613	0.2599	0.2522	0.5115	0.0451	-0.1576	-0.0664	-0.4140	-0.2374	-0.4068	-0.0629	0.0081	0.1462	5.86%
LD5	0.0350	0.2066	0.2604	0.4641	-0.1240	0.2925	0.0913	0.3515	-0.0681	-0.1479	-0.2301	0.0416	-0.1721	2.13%

Table S1 (A_X) | Correlations between acoustic parameters and linear discriminant functions *across all stimuli* in *Group Words* ($N_{Stimuli}$ = 568)

	Duration	*PeakTime*	*Amp.(dB)*	*PeakAmp.*	*Min.F₀*	*Max.F₀*	*Mean.F₀*	*StDev.F₀*	*Jitter*	*Shimmer*	*Max.HNR*	*Mean HNR*	*StDev. HNR*	*Accounted variance*
LD1	-0.7340	-0.5100	0.2539	-0.3841	0.2016	-0.2215	-0.0211	-0.2643	-0.0255	-0.1240	0.1787	0.5830	0.2787	47.06%
LD2	-0.4403	-0.0851	-0.2736	0.2984	0.0027	-0.3055	-0.0228	-0.3721	-0.2419	-0.2118	-0.4198	-0.3058	-0.4197	24.18%
LD3	0.3628	0.4715	-0.0139	0.3205	0.6340	0.3275	0.5176	-0.1632	-0.6386	-0.4825	0.0516	0.5869	0.2807	20.83%
LD4	-0.2250	0.1731	0.4093	0.0664	0.0171	-0.2319	-0.3869	-0.2720	0.1946	0.3110	-0.2801	-0.2053	-0.2558	5.15%
LD5	0.0568	-0.1988	0.2449	0.1941	-0.5855	-0.2789	-0.6499	-0.1307	-0.0310	-0.2910	0.0567	0.2516	-0.0732	2.77%

Table S1 (A_{XI}) | Correlations between acoustic parameters and linear discriminant functions *across all stimuli* in *Group Sentences* ($N_{Stimuli}$ = 470)

	Duration	*PeakTime*	*Amp.(dB)*	*PeakAmp.*	*Min.F₀*	*Max.F₀*	*Mean.F₀*	*StDev.F₀*	*Jitter*	*Shimmer*	*Max.HNR*	*Mean HNR*	*StDev.HNR*	*Accounted variance*
LD1	0.2454	-0.1582	0.0689	0.0624	-0.4932	-0.2370	-0.7993	-0.2683	0.4566	0.0072	-0.1686	0.2165	0.1880	61.42%
LD2	0.1610	0.3348	-0.1514	0.0262	0.4616	0.2975	0.2903	0.1097	0.0316	0.1914	0.1307	0.1549	0.5191	14.99%
LD3	-0.3548	-0.4908	0.5938	-0.1484	0.4706	0.0903	0.1770	-0.0344	-0.2090	-0.2073	-0.1103	0.0541	0.0087	9.85%
LD4	0.2757	0.3209	0.2310	0.0494	0.0690	-0.3130	-0.1124	-0.5473	-0.3181	-0.4536	-0.2842	0.0259	0.2102	7.03%
LD5	0.1057	0.0898	0.2623	0.3306	-0.2243	0.4989	0.0110	0.6534	0.1939	-0.1276	-0.2581	-0.1858	0.1575	3.85%
LD6	0.4486	0.12639	-0.2887	0.6228	-0.1213	0.0009	-0.2277	0.0947	0.3864	0.5177	0.3495	0.4710	-0.4317	2.86%

LDA: The results of all linear discriminant functions and the accounted variance for each function

Table S1 (A_X) | Correlations between acoustic parameters and linear discriminant functions *across all stimuli* in *Group Words* ($N_{Stimuli}$ = 568)

	Duration	*PeakTime*	*Amp.(dB)*	*PeakAmp.*	*Min.F_0*	*Max.F_0*	*Mean.F_0*	*StDev.F_0*	*Jitter*	*Shimmer*	*Max.HNR*	*Mean HNR*	*StDev. HNR*	*Accounted variance*
LD1	-0.7340	-0.5100	0.2539	-0.3841	0.2016	-0.2215	-0.0211	-0.2643	-0.0255	-0.1240	0.1787	0.5830	0.2787	47.06%
LD2	-0.4403	-0.0851	-0.2736	0.2984	0.0027	-0.3055	-0.0228	-0.3721	-0.2419	-0.2118	-0.4198	-0.3058	-0.4197	24.18%
LD3	0.3628	0.4715	-0.0139	0.3205	0.6340	0.3275	0.5176	-0.1632	-0.6386	-0.4825	0.0516	0.5869	0.2807	20.83%
LD4	-0.2250	0.1731	0.4093	0.0664	0.0171	-0.2319	-0.3869	-0.2720	0.1946	0.3110	-0.2801	-0.2053	-0.2558	5.15%
LD5	0.0568	-0.1988	0.2449	0.1941	-0.5855	-0.2789	-0.6499	-0.1307	-0.0310	-0.2910	0.0567	0.2516	-0.0732	2.77%

Table S1 (A_{X1}) | Correlations between acoustic parameters and linear discriminant functions *across all stimuli* in *Group Sentences* ($N_{Stimuli}$ = 470)

	Duration	*PeakTime*	*Amp.(dB)*	*PeakAmp.*	*Min.F_0*	*Max.F_0*	*Mean.F_0*	*StDev.F_0*	*Jitter*	*Shimmer*	*Max.HNR*	*Mean HNR*	*StDev.HNR*	*Accounted variance*
LD1	0.2454	-0.1582	0.0689	0.0624	-0.4932	-0.2370	-0.7993	-0.2683	0.4566	0.0072	-0.1686	0.2165	0.1880	61.42%
LD2	0.1610	0.3348	-0.1514	0.0262	0.4616	0.2975	0.2903	0.1097	0.0316	0.1914	0.1307	0.1549	0.5191	14.99%
LD3	-0.3548	-0.4908	0.5938	-0.1484	0.4706	0.0903	0.1770	-0.0344	-0.2090	-0.2073	-0.1103	0.0541	0.0087	9.85%
LD4	0.2757	0.3209	0.2310	0.0494	0.0690	-0.3130	-0.1124	-0.5473	-0.3181	-0.4536	-0.2842	0.0259	0.2102	7.03%
LD5	0.1057	0.0898	0.2623	0.3306	-0.2243	0.4989	0.0110	0.6534	0.1939	-0.1276	-0.2581	-0.1858	0.1575	3.85%
LD6	0.4486	0.12639	-0.2887	0.6228	-0.1213	0.0009	-0.2277	0.0947	0.3864	0.5177	0.3495	0.4710	-0.4317	2.86%

Table S1 (A_{X2}) | Correlations between acoustic parameters and linear discriminant functions *across all stimuli* ($N_{Stimuli}$ = 1038)

	Duration	*PeakTime*	*Amp.(dB)*	*PeakAmp.*	*Min.F_0*	*Max.F_0*	*Mean.F_0*	*StDev.F_0*	*Jitter*	*Shimmer*	*Max.HNR*	*Mean HNR*	*StDev.HNR*	*Accounted variance*
LD1	-0.0409	-0.1940	0.2412	-0.2334	-0.0884	-0.1356	-0.4198	-0.1066	0.2979	0.0862	0.1062	0.4662	0.3770	45.03%
LD2	-0.0424	-0.2201	-0.1633	-0.0019	-0.8696	-0.3892	-0.7815	0.0701	0.5785	0.3505	-0.2404	-0.5629	-0.4285	24.23%
LD3	0.5136	0.5155	0.0075	-0.0094	-0.1365	0.5270	0.2164	0.6458	0.3064	0.3170	0.3913	-0.0529	0.3650	13.97%
LD4	-0.5356	-0.3553	-0.3182	-0.4923	0.0134	-0.2441	0.1541	0.0080	0.2375	0.2505	0.0968	-0.0183	-0.0548	8.69%
LD5	0.0975	-0.0251	0.8187	-0.1977	-0.1924	0.1579	-0.0070	0.0302	0.0769	-0.1269	0.1785	0.1147	-0.0512	4.68%
LD6	0.0804	-0.2092	-0.1228	0.2788	-0.1775	0.2978	0.0995	0.3782	-0.1606	-0.3977	0.3855	0.3846	0.2249	3.40%

Figure S1

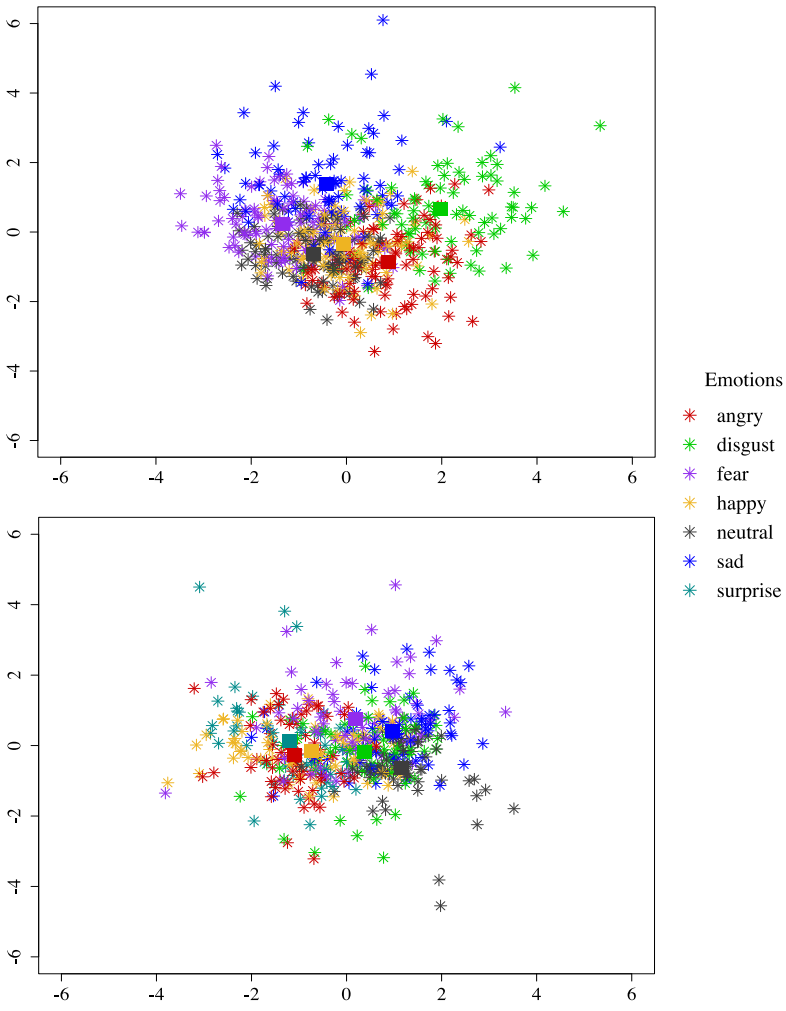


Figure S1 | Linear discriminant analysis: **(a)** Across all stimuli types ($N_{stimuli}$ = 568) presented in *Group Words*. Each stimulus is plotted according to its scores for the discriminant function 1 (strongest correlation with *Duration, Mean HNR & Peak time)* and function 2 (strongest correlation with *Duration, Maximum HNR & Standard deviation HNR*). **(b)** Across all stimuli types ($N_{stimuli}$ = 470) presented in *Group Sentences*. Each stimulus is plotted according to its scores for the discriminant function 1 (strongest correlation with *Mean F_0, Minimum F_0 & Jitter)* and function 2 (strongest correlation with *Standard deviation HNR, Minimum F_0 & Peak time*). On the ***x-axis*** is the *linear discriminant function 1* displayed, while on the ***y-axis*** the *linear discriminant function 2*. The squares represent the means of the emotion groups on each of the discriminant functions.

Figure S2A

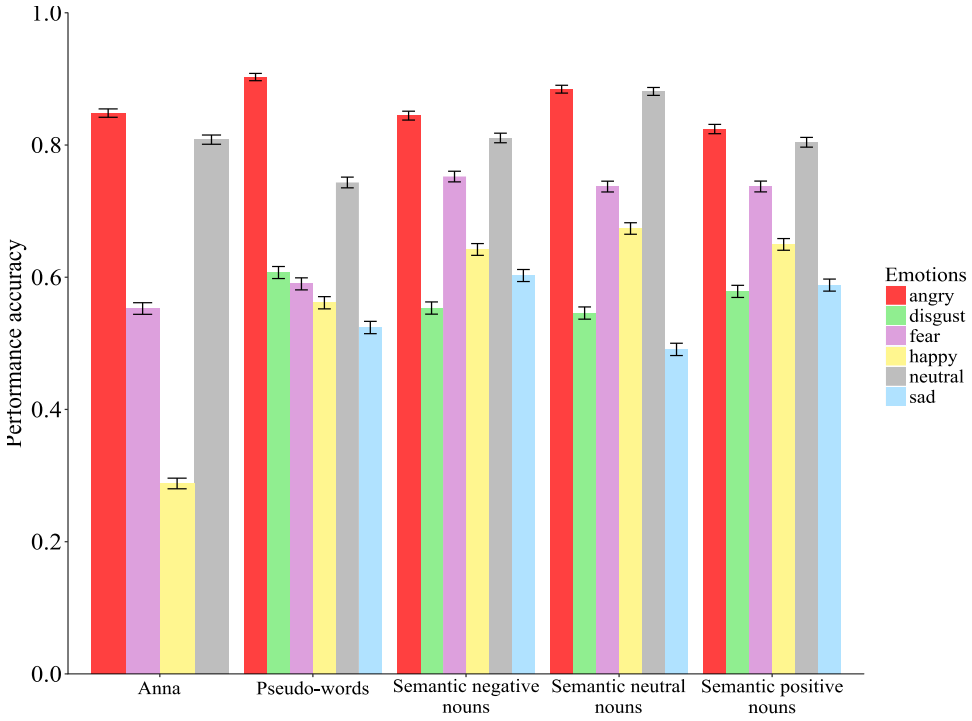


FIGURE S2 | (A) *Group Words* (*n* = 145). The bars represent raters' accuracy for each emotion category. The rate of correct random recognition assuming a uniform probability distribution was 1/7 (14%). As it can be observed listeners recognized all emotions at rates that were significantly higher than random performance.

Figure S2B

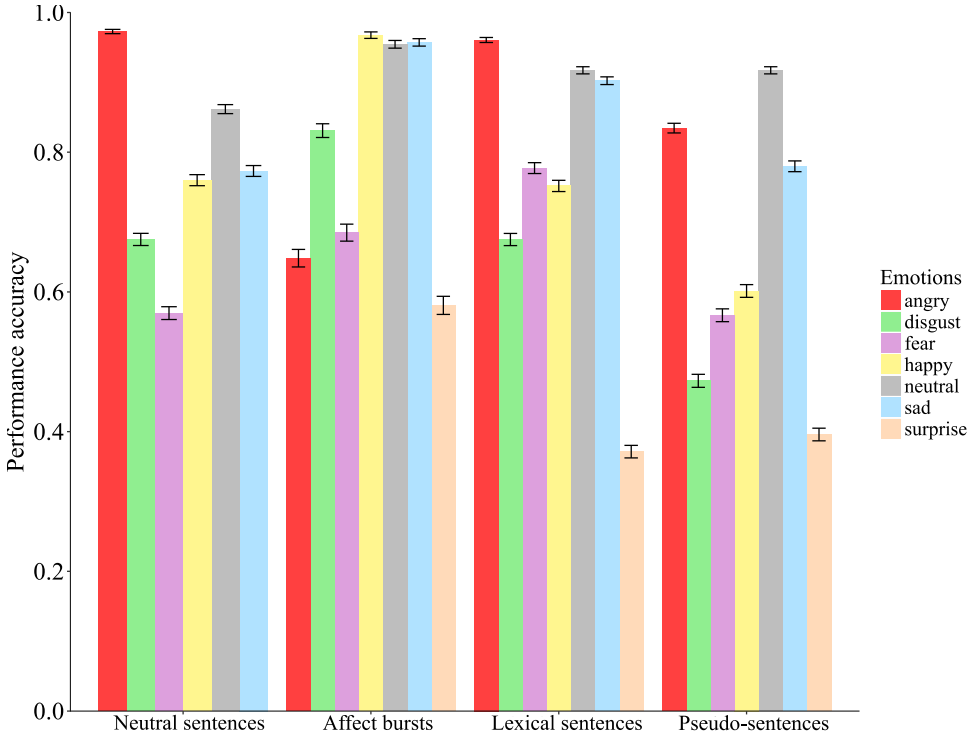


FIGURE S2 | (B) *Group Sentences* (*n* = 145). The bars represent raters' accuracy for each emotion category. The rate of correct random recognition assuming a uniform probability distribution was 1/7 (14%). As it can be observed listeners recognized all emotions at rates that were significantly higher than random performance.

Emotion recognition and confidence ratings across all stimuli

Global models – Group Words & Group Sentences

Table S2 (A) | *Quasi-binomial logistic model across all stimuli types and emotion categories in* ***Group Words*** (***DV*** = **Emotion Recognition**)

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			82051	103009	
Participants	144	2014.1	81907	100995	< .001
Duration	1	564.0	81906	100431	< .001
Standard deviation HNR	1	257.9	81905	100173	< .001
Peak time	1	210.7	81904	99962	< .001
Mean HNR	1	41.4	81903	99921	< .001
Shimmer	1	82.4	81902	99839	< .001
Mean F_0	1	34.1	81901	99805	< .001
Maximum F_0	1	41.4	81900	99763	< .001
Amplitude (dB.)	1	16.1	81899	99747	< .001
Jitter	1	14.1	81898	99733	< .001
Minimum F_0	1	7.8	81897	99725	.006
Stimuli types	4	936.0	81893	98789	< .001
Emotions	5	5376.6	81888	93413	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Table S2 (A_I) | *Linear model across all stimuli types and emotion categories in* ***Group Words*** (***DV*** = **Confidence ratings**)

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			82051	200359	
Participants	144	40150	81907	160210	< .001
Duration	1	568	81906	159641	< .001
Standard deviation HNR	1	357	81905	159285	< .001
Peak time	1	854	81904	158431	< .001
Mean HNR	1	100	81903	158331	< .001
Shimmer	1	449	81902	157882	< .001
Mean F_0	1	686	81901	157196	< .001
Maximum F_0	1	101	81900	157095	< .001
Amplitude (dB.)	1	19	81899	157076	.001
Jitter	1	49	81898	157027	< .001
Minimum F_0	1	58	81897	156969	< .001
Stimuli types	4	865	81893	156104	< .001
Emotions	5	5721	81888	150383	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Table S2 (A_{II}) | *Linear model across all stimuli types and emotion categories in* ***Group Words*** (**Confidence predicted by correct emotion recognition**)

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			82051	200359	
Participants	144	40150	81907	160210	< .001
Stimuli types	4	450	81903	159759	< .001
Correct emotion recognition	1	12770	81902	146990	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Table S2 (B) | *Quasi-binomial logistic model across all stimuli types in* ***Group Sentences*** (***DV*** = **Emotion Recognition**)

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			68138	78349	
Participants	144	794.4	67994	77554	< .001
Amplitude (dB.)	1	487.5	67993	77067	< .001
Mean HNR	1	218.5	67992	76848	< .001
Duration	1	143.1	67991	76705	< .001
Minimum F_0	1	119.4	67990	76586	< .001
Peak time	1	42.3	67989	76544	< .001
Peak Amplitude (dB.)	1	24.8	67988	76519	< .001
Jitter	1	17.1	67987	76502	< .001
Maximum F_0	1	18.1	67986	76484	< .001
Standard deviation F_0	1	13.8	67985	76470	< .001
Maximum HNR	1	3.7	67984	76466	.059
Shimmer	1	4.4	67983	76462	.040
Stimuli types	3	1379.3	67980	75082	< .001
Emotions	6	8160.0	67974	66923	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Table S2 (B_I) | *Linear model across all stimuli types in* ***Group Sentences*** (***DV*** = **Confidence ratings**)

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			68135	173140	
Participants	144	32748	67991	140392	< .001
Amplitude (dB.)	1	773	67990	139618	< .001
Mean HNR	1	79	67989	139539	< .001
Duration	1	782	67988	138757	< .001
Minimum F_0	1	18	67987	138739	.002
Peak time	1	1	67986	138738	.487
Peak Amplitude (dB.)	1	118	67985	138620	< .001
Jitter	1	701	67984	137919	< .001
Maximum F_0	1	395	67983	137524	< .001
Standard deviation F_0	1	473	67982	137052	< .001
Maximum HNR	1	84	67981	136968	< .001
Shimmer	1	15	67980	136953	.005
Stimuli types	3	3502	67977	133451	< .001
Emotions	6	6593	67971	126858	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Table S2 (B_{II}) | *Linear model across all stimuli types in* ***Group Sentences*** (**Confidence predicted by correct emotion recognition**)

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			68124	173113	
Participants	144	32746	67980	140367	< .001
Stimuli types	3	3208	67977	137159	< .001
Correct emotion recognition	1	11792	67976	125367	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Emotion recognition and confidence ratings across all stimuli

Conditional models

Table S2 (C) | *Quasi-binomial logistic model for* ***Anna*** **(*DV* = Emotion Recognition)**

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			12756	16889	
Participants	144	211.31	12612	16678	< .001
Standard deviation HNR	1	351.39	12611	16327	< .001
Maximum F_0	1	122.65	12610	16204	< .001
Minimum F_0	1	307.95	12609	15896	< .001
Amplitude (dB.)	1	191.40	12608	15705	< .001
Duration	1	181.75	12607	15523	< .001
Maximum HNR	1	90.29	12606	15433	< .001
Standard deviation F0	1	15.92	12605	15417	< .001
Peak amplitude	1	16.32	12604	15400	< .001
Jitter	1	8.01	12603	15392	.005
Mean HNR	1	7.09	12602	15385	.008
Peak time	1	6.56	12601	15379	.011
Emotions	3	1911.39	12598	13467	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Table S2 (C_I) | *Linear model for* ***Anna*** **(*DV* = Confidence ratings)**

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			12756	30714	
Participants	144	6232.4	12612	24482	< .001
Standard deviation HNR	1	440.0	12611	24042	< .001
Maximum F_0	1	99.5	12610	23942	< .001
Minimum F_0	1	124.5	12609	23818	< .001
Amplitude (dB.)	1	236.3	12608	23582	< .001
Duration	1	1.6	12607	23580	.346
Maximum HNR	1	26.6	12606	23553	< .001
Standard deviation F0	1	37.3	12605	23516	< .001
Peak amplitude	1	0.1	12604	23516	.856
Jitter	1	41.6	12603	23474	< .001
Mean HNR	1	1.1	12602	23473	.431
Peak time	1	1.2	12601	23472	.412
Emotions	3	768.1	12598	22704	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Table S2 (C_{II}) | *Linear model for* ***Anna*** **(Confidence predicted by correct emotion recognition)**

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			12756	30714	
Participants	144	6232.4	12612	24482	< .001
Correct emotion recognition	1	2742.5	12611	21739	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Table S2 (D) | *Quasi-binomial logistic model for* ***Pseudo-words*** **(*DV* = Emotion Recognition)**

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			17104	22038	
Participants	144	1037.94	16960	21001	< .001
Maximum HNR	1	331.93	16959	20669	< .001
Mean F_0	1	161.64	16958	20507	< .001
Shimmer	1	74.70	16957	20432	< .001
Mean HNR	1	98.49	16956	20334	< .001
Peak amplitude (dB.)	1	96.53	16955	20237	< .001
Standard deviation HNR	1	51.27	16954	20186	< .001
Maximum F_0	1	46.54	16953	20140	< .001
Standard deviation F_0	1	12.64	16952	20127	< .001
Minimum F_0	1	5.79	16951	20121	.018
Peak time	1	2.32	16950	20119	.134
Duration	1	4.75	16949	20114	.032
Emotions	5	965.77	16944	19148	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Table S2 (D_I) | *Linear model for* ***Pseudo-words*** **(*DV* = Confidence ratings)**

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			17104	41694	
Participants	144	9501.9	16960	32192	< .001
Maximum HNR	1	456.2	16959	31736	< .001
Mean F_0	1	645.7	16958	31090	< .001
Shimmer	1	384.2	16957	30706	< .001
Mean HNR	1	171.2	16956	30535	< .001
Peak amplitude (dB.)	1	10.4	16955	30524	.014
Standard deviation HNR	1	69.3	16954	30455	< .001
Maximum F_0	1	58.7	16953	30396	< .001
Standard deviation F_0	1	29.5	16952	30367	< .001
Minimum F_0	1	55.6	16951	30311	< .001
Peak time	1	19.1	16950	30292	< .001
Duration	1	110.9	16949	30181	< .001
Emotions	5	1201.1	16944	28980	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Table S2 (D_{II}) | *Linear model for* ***Pseudo-words*** **(Confidence predicted by correct emotion recognition)**

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			17104	41694	
Participants	144	9501.9	16960	32192	< .001
Correct emotion recognition	1	2196.2	16959	29995	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Emotion recognition and confidence ratings across all stimuli

Conditional models

Table S2 (E) | *Quasi-binomial logistic model for* ***Semantic positive nouns*** (***DV*** = **Emotion Recognition**)

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			17396	21343	
Participants	144	643.12	17252	20700	< .001
Shimmer	1	91.30	17251	20609	< .001
Jitter	1	216.90	17250	20392	< .001
Duration	1	158.73	17249	20233	< .001
Standard deviation HNR	1	90.98	17248	20142	< .001
Amplitude (dB.)	1	33.70	17247	20109	< .001
Mean F_0	1	36.16	17246	20072	< .001
Minimum F_0	1	8.38	17245	20064	.004
Mean HNR	1	4.85	17244	20059	.031
Maximum F_0	1	3.97	17243	20055	.050
Peak amplitude (dB.)	1	2.73	17242	20052	.105
Emotions	5	617.03	17237	19435	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Table S2 (E_I) | *Linear model for* ***Semantic positive nouns*** (***DV*** = **Confidence ratings**)

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			17396	42645	
Participants	144	9458.2	17252	33187	< .001
Shimmer	1	40.8	17251	33146	< .001
Jitter	1	292.3	17250	32854	< .001
Duration	1	251.0	17249	32603	< .001
Standard deviation HNR	1	97.2	17248	32506	< .001
Amplitude (dB.)	1	3.6	17247	32502	.160
Mean F_0	1	37.6	17246	32465	< .001
Minimum F_0	1	49.5	17245	32415	< .001
Mean HNR	1	19.1	17244	32396	.001
Maximum F_0	1	88.5	17243	32308	< .001
Peak amplitude (dB.)	1	27.2	17242	32280	< .001
Emotions	5	869.5	17237	31411	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Table S2 (E_{II}) | *Linear model for* ***Semantic positive nouns*** (**Confidence predicted by correct emotion recognition**)

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			17396	42645	
Participants	144	9458.2	17252	33187	< .001
Correct emotion recognition	1	2367.0	17251	30820	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Table S2 (F) | *Quasi-binomial logistic model for* ***Semantic negative nouns*** (***DV*** = **Emotion Recognition**)

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			17395	21225	
Participants	144	604.10	17251	20620	< .001
Duration	1	196.75	17250	20424	< .001
Peak time	1	126.72	17249	20297	< .001
Amplitude (dB.)	1	81.93	17248	20215	< .001
Maximum F_0	1	60.12	17247	20155	< .001
Mean HNR	1	30.09	17246	20125	< .001
Standard deviation HNR	1	44.05	17245	20081	< .001
Standard deviation F_0	1	15.90	17244	20065	< .001
Peak amplitude (dB.)	1	16.84	17243	20048	< .001
Shimmer	1	11.91	17242	20036	< .001
Jitter	1	14.23	17241	20022	< .001
Maximum HNR	1	8.03	17240	20014	.005
Emotions	5	792.14	17235	19222	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Table S2 (F_I) | *Linear model for* ***Semantic negative nouns*** (***DV*** = **Confidence ratings**)

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			17395	41929	
Participants	144	9368.3	17251	32560	< .001
Duration	1	224.1	17250	32336	< .001
Peak time	1	186.4	17249	32150	< .001
Amplitude (dB.)	1	216.3	17248	31933	< .001
Maximum F_0	1	100.6	17247	31833	< .001
Mean HNR	1	7.5	17246	31825	.039
Standard deviation HNR	1	11.7	17245	31814	.010
Standard deviation F_0	1	12.6	17244	31801	.007
Peak amplitude (dB.)	1	131.0	17243	31670	< .001
Shimmer	1	64.0	17242	31606	< .001
Jitter	1	37.5	17241	31569	< .001
Maximum HNR	1	80.2	17240	31488	< .001
Emotions	5	1236.0	17235	30252	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Table S2 (F_{II}) | *Linear model for* ***Semantic negative nouns*** (**Confidence predicted by correct emotion recognition**)

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			17395	41929	
Participants	144	9368.3	17251	32560	< .001
Correct emotion recognition	1	2688.5	17250	29872	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Emotion recognition and confidence ratings across all stimuli

Conditional models

Table S2 (G) | *Quasi-binomial logistic model for* ***Semantic neutral nouns*** (***DV*** = **Emotion Recognition**)

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			17396	21190	
Participants	144	675.44	17252	20514	< .001
Duration	1	275.92	17251	20238	< .001
Peak amplitude (dB.)	1	301.38	17250	19937	< .001
Amplitude (dB.)	1	247.23	17249	19690	< .001
Jitter	1	172.25	17248	19517	< .001
Mean HNR	1	115.07	17247	19402	< .001
Shimmer	1	127.09	17246	19275	< .001
Peak time	1	119.92	17245	19155	< .001
Standard deviation F_0	1	9.89	17244	19145	.002
Minimum F_0	1	4.71	17243	19141	.034
Mean F_0	1	21.12	17242	19119	< .001
Standard deviation HNR	1	16.62	17241	19103	< .001
Maximum HNR	1	7.21	17240	19096	.009
Emotions	5	1108.37	17235	17987	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Table S2 (G_1) | *Linear model for* ***Semantic neutral nouns*** (***DV*** = **Confidence ratings**)

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			17396	42927	
Participants	144	9383.4	17252	33543	< .001
Duration	1	315.4	17251	33228	< .001
Peak amplitude (dB.)	1	429.2	17250	32799	< .001
Amplitude (dB.)	1	183.6	17249	32615	< .001
Jitter	1	183.9	17248	32431	< .001
Mean HNR	1	127.5	17247	32304	< .001
Shimmer	1	171.5	17246	32132	< .001
Peak time	1	148.4	17245	31984	< .001
Standard deviation F_0	1	4.1	17244	31980	.129
Minimum F_0	1	9.2	17243	31971	.024
Mean F_0	1	0.3	17242	31970	.669
Standard deviation HNR	1	33.5	17241	31937	< .001
Maximum HNR	1	49.6	17240	31887	< .001
Emotions	5	1081.7	17235	30806	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Table S2 (G_{II}) | *Linear model for* ***Semantic neutral nouns*** (**Confidence predicted by correct emotion recognition**)

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			17396	42927	
Participants	144	9383.4	17252	33543	< .001
Correct emotion recognition	1	2576.3	17251	30967	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Table S2 (H) | *Quasi-binomial logistic model for* ***Affect bursts*** (***DV*** = **Emotion Recognition**)

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			10147	10059.5	
Participants	144	174.82	10003	9884.7	.069
Peak amplitude (dB.)	1	352.53	10002	9532.1	< .001
Mean HNR	1	678.68	10001	8853.5	< .001
Standard deviation HNR	1	176.93	10000	8676.5	< .001
Peak time	1	88.31	9999	8588.2	< .001
Maximum HNR	1	57.98	9998	8530.2	< .001
Shimmer	1	9.86	9997	8520.4	.002
Mean F_0	1	13.83	9996	8506.6	< .001
Standard deviation F_0	1	12.91	9995	8493.7	< .001
Amplitude (dB.)	1	9.31	9994	8484.3	.003
Minimum F_0	1	3.85	9993	8480.5	.053
Duration	1	2.84	9992	8477.7	.097
Emotions	6	684.60	9986	7793.1	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Table S2 (H_1) | *Linear model for* ***Affect bursts*** (***DV*** = **Confidence ratings**)

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			10148	24651	
Participants	144	6092.5	10004	18558	< .001
Peak amplitude (dB.)	1	788.8	10003	17770	< .001
Mean HNR	1	175.8	10002	17594	< .001
Standard deviation HNR	1	260.3	10001	17334	< .001
Peak time	1	122.4	10000	17211	< .001
Maximum HNR	1	95.9	9999	17115	< .001
Shimmer	1	35.9	9998	17079	< .001
Mean F_0	1	4.6	9997	17075	.091
Standard deviation F_0	1	42.1	9996	17033	< .001
Amplitude (dB.)	1	23.9	9995	17009	< .001
Minimum F_0	1	95.9	9994	16913	< .001
Duration	1	103.5	9993	16809	< .001
Emotions	6	676.4	9987	16133	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Table S2 (H_{II}) | *Linear model for* ***Affect bursts*** (**Confidence predicted by correct emotion recognition**)

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			10146	24649	
Participants	144	6091.1	10002	18558	< .001
Correct emotion recognition	1	2644.1	10001	15914	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Emotion recognition and confidence ratings across all stimuli

Conditional models

Table S2 (I) | *Quasi-binomial logistic model for* ***Pseudo-sentences*** **(*DV* = Emotion Recognition)**

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			20295	26215	
Participants	144	391.64	20151	25824	< .001
Standard deviation HNR	1	145.09	20150	25679	< .001
Duration	1	223.10	20149	25456	< .001
Minimum F_0	1	214.47	20148	25241	< .001
Peak time	1	73.02	20147	25168	< .001
Amplitude (dB.)	1	54.94	20146	25113	< .001
Standard deviation F_0	1	33.70	20145	25080	< .001
Mean F_0	1	39.94	20144	25040	< .001
Maximum HNR	1	26.12	20143	25013	< .001
Peak amplitude (dB.)	1	17.28	20142	24996	< .001
Shimmer	1	8.84	20141	24987	.003
Mean HNR	1	3.28	20140	24984	.069
Jitter	1	7.58	20139	24976	.006
Emotions	6	3015.30	20133	21961	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Table S2 (I_I) | *Linear model for* ***Pseudo-sentences*** **(*DV* = Confidence ratings)**

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			20296	56886	
Participants	144	13738.8	20152	43147	< .001
Standard deviation HNR	1	15.1	20151	43132	.005
Duration	1	379.3	20150	42752	< .001
Minimum F_0	1	0.2	20149	42752	.753
Peak time	1	102.3	20148	42650	< .001
Amplitude (dB.)	1	171.0	20147	42479	< .001
Standard deviation F_0	1	843.8	20146	41635	< .001
Mean F_0	1	363.3	20145	41272	< .001
Maximum HNR	1	29.7	20144	41242	< .001
Peak amplitude (dB.)	1	110.1	20143	41132	< .001
Shimmer	1	48.2	20142	41084	< .001
Mean HNR	1	10.6	20141	41073	.018
Jitter	1	40.8	20140	41033	< .001
Emotions	6	3011.0	20134	38022	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Table S2 (I_{II}) | *Linear model for* ***Pseudo-sentences*** **(Confidence predicted by correct emotion recognition)**

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			20292	56881	
Participants	144	13735.7	20148	43145	< .001
Correct emotion recognition	1	3410.1	20147	39735	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Table S2 (J) | *Quasi-binomial logistic model for* ***Lexical sentences*** **(*DV* = Emotion Recognition)**

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			20295	22129	
Participants	144	452.6	20151	21676	< .001
Duration	1	425.4	20150	21251	< .001
Standard deviation F_0	1	349.6	20149	20901	< .001
Amplitude (dB.)	1	51.9	20148	20849	< .001
Peak amplitude (dB.)	1	36.8	20147	20813	< .001
Mean F_0	1	26.9	20146	20786	< .001
Shimmer	1	40.5	20145	20745	< .001
Minimum F_0	1	11.6	20144	20734	< .001
Maximum HNR	1	11.2	20143	20722	< .001
Peak time	1	8.5	20142	20714	.004
Maximum F_0	1	7.6	20141	20706	.006
Jitter	1	5.7	20140	20701	.018
Emotions	6	3531.2	20134	17169	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Table S2 (J_I) | *Linear model for* ***Lexical sentences*** **(*DV* = Confidence ratings)**

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			20295	44310	
Participants	144	10933.2	20151	33377	< .001
Duration	1	113.3	20150	33263	< .001
Standard deviation F_0	1	94.2	20149	33169	< .001
Amplitude (dB.)	1	111.2	20148	33058	< .001
Peak amplitude (dB.)	1	144.0	20147	32914	< .001
Mean F_0	1	1058.1	20146	31856	< .001
Shimmer	1	166.9	20145	31689	< .001
Minimum F_0	1	326.7	20144	31362	< .001
Maximum HNR	1	49.5	20143	31313	< .001
Peak time	1	40.3	20142	31273	< .001
Maximum F_0	1	16.4	20141	31256	< .001
Jitter	1	96.0	20140	31160	< .001
Emotions	6	1647.8	20134	29512	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Table S2 (J_{II}) | *Linear model for* ***Lexical sentences*** **(Confidence predicted by correct emotion recognition)**

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			20291	44294	
Participants	144	10934.6	20147	33360	< .001
Correct emotion recognition	1	1728.7	20146	31631	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Emotion recognition and confidence ratings across all stimuli

Conditional models

Table S2 (K) | *Quasi-binomial logistic model for* ***Neutral sentences*** **(*DV* = Emotion Recognition)**

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			17398	18819	
Participants	144	568.43	17254	18251	< .001
Amplitude (dB.)	1	227.36	17253	18024	< .001
Shimmer	1	58.28	17252	17965	< .001
Standard deviation F_0	1	77.14	17251	17888	< .001
Maximum HNR	1	28.06	17250	17860	< .001
Mean HNR	1	46.80	17249	17813	< .001
Duration	1	16.29	17248	17797	< .001
Peak time	1	31.74	17247	17765	< .001
Minimum F_0	1	11.65	17246	17754	< .001
Jitter	1	19.32	17245	17734	< .001
Peak amplitude (dB.)	1	12.69	17244	17722	< .001
Standard deviation HNR	1	8.18	17243	17713	.007
Emotions	5	2079.31	17238	15634	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Table S2 (K_I) | *Linear model for* ***Neutral sentences*** **(*DV* = Confidence ratings)**

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			17398	18819	
Participants	144	9105.0	17249	34981	< .001
Amplitude (dB.)	1	764.3	17248	34217	< .001
Shimmer	1	13.4	17247	34203	.007
Standard deviation F_0	1	176.0	17246	34027	< .001
Maximum HNR	1	245.2	17245	33782	< .001
Mean HNR	1	303.0	17244	33479	< .001
Duration	1	2.0	17243	33477	.289
Peak time	1	5.3	17242	33472	.086
Minimum F_0	1	9.4	17241	33462	.023
Jitter	1	5.9	17240	33456	.070
Peak amplitude (dB.)	1	1.1	17239	33455	.438
Standard deviation HNR	1	23.1	17238	33432	< .001
Emotions	5	2223.5	17233	31209	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Table S2 (K_{II}) | *Linear model for* ***Neutral sentences*** **(Confidence predicted by correct emotion recognition)**

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*
NULL			17392	44080	
Participants	144	9102.5	17248	34977	< .001
Correct emotion recognition	1	4250.1	17247	30727	< .001

Note: *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance.

Odds ratio (OR) & linear contrasts (Δ) for emotion comparisons

Table S3 | *Odds ratio* estimates for recognition accuracy and *linear contrasts* for the pattern of the differences in the expressed confidence for the comparisons between emotion categories in *Group Words* (*N* = 145) and *Group Sentences* (*N* =145)

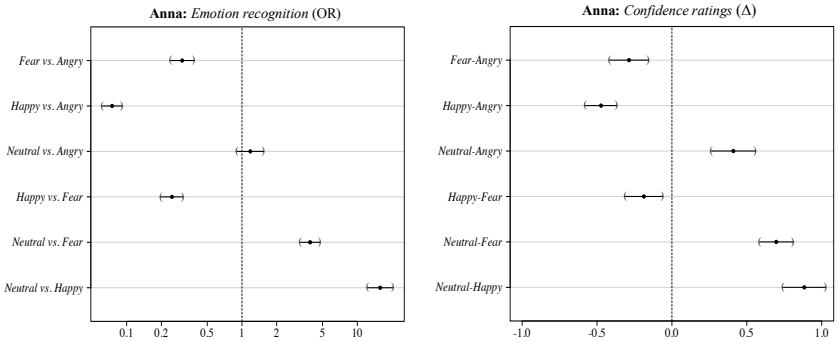
Emotion comparisons	**Global**			
	Group Words		**Group Sentences**	
	Emotion recognition	***Confidence ratings***	***Emotion recognition***	***Confidence ratings***
	Odds ratio (OR)	*Difference (Δ)*	*Odds ratio (OR)*	*Difference (Δ)*
Disgust vs. Angry	0.19, [0.17; 0.21]	-0.84, [-0.90; -0.79]	0.24, [0.21; 0.27]	-0.83, [-0.89; -0.76]
Fear vs. Angry	0.36, [0.32; 0.40]	-0.44, [-0.50; -0.39]	0.24, [0.21; 0.27]	-0.90, [-0.96; -0.84]
Happy vs. Angry	0.20, [0.18; 0.22]	-0.48, [-0.53; -0.43]	0.36, [0.32; 0.40]	-0.49, [-0.55; -0.43]
Neutral vs. Angry	0.77, [0.69; 0.85]	-0.18, [-0.23; -0.13]	1.80, [1.55; 2.09]	-0.07, [-0.13; -0.00]
Sad vs. Angry	0.19, [0.17; 0.21]	-0.88, [-0.94; -0.83]	0.83, [0.73; 0.95]	-0.42, [-0.48; -0.35]
Surprise vs. Angry	–	–	0.08, [0.07; 0.09]	-0.55, [-0.61; -0.48]
Fear vs. Disgust	1.87, [1.68; 2.09]	0.40, [0.33; 0.46]	0.98, [0.89; 1.08]	-0.08, [-0.14; -0.02]
Happy vs. Disgust	1.04, [0.94; 1.14]	0.36, [0.30; 0.42]	1.49, [1.34; 1.65]	0.34, [0.28; 0.40]
Neutral vs. Disgust	4.01, [3.58; 4.49]	0.66, [0.60; 0.73]	7.47, [6.58; 8.48]	0.76, [0.70; 0.82]
Sad vs. Disgust	1.00, [0.90; 1.10]	-0.04, [-0.10; 0.02]	3.45, [3.09; 3.85]	0.41, [0.35; 0.47]
Surprise vs. Disgust	–	–	0.33, [0.29; 0.37]	0.28, [0.21; 0.35]
Happy vs. Fear	0.55, [0.51; 0.60]	-0.04, [-0.09; 0.02]	1.52, [1.38; 1.68]	0.41, [0.35; 0.47]
Neutral vs. Fear	2.14, [1.95; 2.35]	0.26, [0.21; 0.32]	7.65, [6.72; 8.71]	0.83, [0.77; 0.89]
Sad vs. Fear	0.53, [0.49; 0.59]	-0.44, [-0.50; -0.38]	3.53, [3.15; 3.95]	0.48, [0.42; 0.54]
Surprise vs. Fear	–	–	0.33, [0.30; 0.37]	0.35, [0.29; 0.42]
Neutral vs. Happy	3.87, [3.55; 4.21]	0.30, [0.25; 0.35]	5.02, [4.39; 5.74]	0.42, [0.36; 0.48]
Sad vs. Happy	0.96, [0.88; 1.05]	-0.40, [-0.45; -0.35]	2.32, [2.06; 2.61]	0.07, [0.01; 0.13]
Surprise vs. Happy	–	–	0.22, [0.20; 0.24]	-0.06, [-0.12; 0.00]
Sad vs. Neutral	0.25, [0.23; 0.27]	-0.70, [-0.75; -0.65]	0.46, [0.40; 0.53]	-0.35, [-0.41; -0.29]
Surprise vs. Neutral	–	–	0.04, [0.04; 0.05]	-0.48, [-0.55; -0.41]
Surprise vs. Sad	–	–	0.09, [0.08; 0.11]	-0.13, [-0.20; -0.06]

Note: In the squared brackets are *Tukey 95% confidence intervals* displayed. Odds ratio of emotion 1 (e.g., disgust) vs. emotion 2 (e.g., angry) less than 1 indicate that the recognition probability of emotion 2 (e.g., angry) is significantly higher than of emotion 1 (e.g., disgust), whereas values greater than 1 vice-versa. If the odds ratio of 1 is covered in the confidence interval, the difference in the recognition probabilities is not significant. Negative differences of confidence ratings of emotion 1 (e.g., disgust) vs. emotion 2 (e.g., angry) indicate that the confidence ratings of emotion 2 (e.g., angry) is significantly higher than of emotion 1 (e.g., disgust), whereas positive differences vice-versa. If the difference of zero is covered in the 95%CI, the difference in the confidence ratings is not significant.

Odds ratio (OR) & linear contrasts (Δ) for emotion comparisons

Table S4(A) | *Odds ratio* estimates for recognition accuracy and *linear contrasts* for the pattern of the differences in the expressed confidence for the comparisons between emotion categories in Anna stimuli

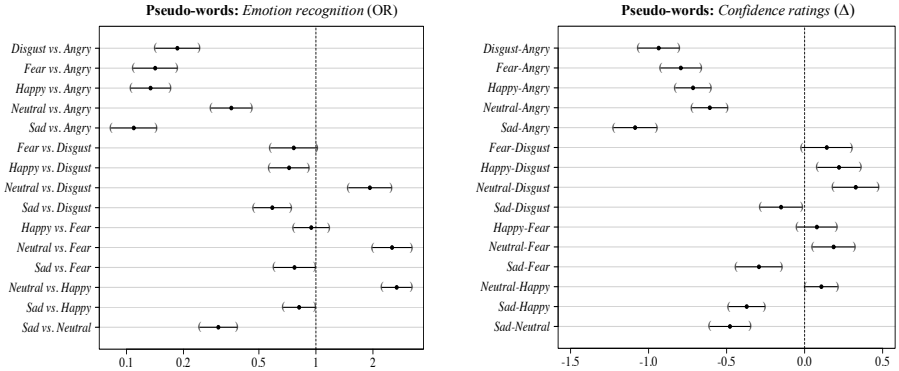
	Anna	
Emotion comparisons	***Emotion recognition***	***Confidence ratings***
	Odds ratio (OR)	*Difference (Δ)*
Fear vs. Angry	0.30, [0.24; 0.38]	-0.29, [-0.41; -0.16]
Happy vs. Angry	0.07, [0.06; 0.09]	-0.47, [-0.58; -0.37]
Neutral vs. Angry	1.18, [0.91; 1.54]	0.41, [0.26; 0.56]
Happy vs. Fear	0.25, [0.20; 0.31]	-0.19, [-0.31; -0.06]
Neutral vs. Fear	3.90, [3.22; 4.73]	0.70, [0.59; 0.81]
Neutral vs. Happy	15.80, [12.31; 20.28]	0.88, [0.74; 1.02]



Note: As it can be observed, listeners were less accurate and less confident at categorizing utterances spoken in a *fear* and *happy* tone of voice than when spoken in an *angry* prosody. Although the recognition accuracy was not significant when comparing *neutral* to *angry* expressions, listeners felt more confident at categorizing Anna stimuli when spoken in a neutral tone of voice. For the other emotion comparisons, the odds of correctly detecting emotions were similar to the pattern of the differences in confidence judgements.

Table S4(B) | *Odds ratio* estimates for recognition accuracy and *linear contrasts* for the pattern of the differences in the expressed confidence for the comparisons between emotion categories in pseudo-words

	Pseudo-words	
Emotion comparisons	**Emotion recognition**	***Confidence ratings***
	Odds ratio (OR)	*Difference (Δ)*
Disgust vs. Angry	0.19, [0.14; 0.24]	-0.94, [-1.07; -0.81]
Fear vs. Angry	0.14, [0.11; 0.18]	-0.79, [-0.92; -0.67]
Happy vs. Angry	0.13, [0.11; 0.17]	-0.72, [-0.83; -0.60]
Neutral vs. Angry	0.36, [0.28; 0.46]	-0.61, [-0.72; -0.50]
Sad vs. Angry	0.11, [0.08; 0.14]	-1.09, [-1.22; -0.95]
Fear vs. Disgust	0.76, [0.58; 1.01]	0.14, [-0.01; 0.30]
Happy vs. Disgust	0.72, [0.57; 0.92]	0.22, [0.08; 0.36]
Neutral vs. Disgust	1.93, [1.49; 2.49]	0.33, [0.18; 0.47]
Sad vs. Disgust	0.59, [0.47; 0.74]	-0.15, [-0.28; -0.02]
Happy vs. Fear	0.95, [0.76; 1.17]	0.08, [-0.05; 0.20]
Neutral vs. Fear	2.52, [1.99; 3.19]	0.19, [0.05; 0.32]
Sad vs. Fear	0.77, [0.60; 0.99]	-0.29, [-0.44; -0.15]
Neutral vs. Happy	2.67, [2.23; 3.19]	0.11, [0.00; 0.21]
Sad vs. Happy	0.82, [0.67; 0.99]	-0.37, [-0.49; -0.26]
Sad vs. Neutral	0.31, [0.24; 0.38]	-0.48, [-0.61; -0.35]

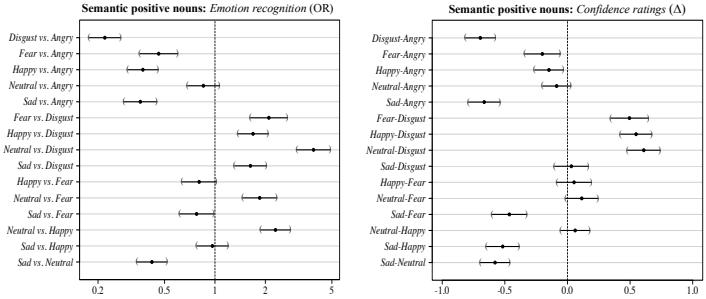


Note: As it can be observed, listeners were significantly less accurate and rated themselves as less confident when identifying emotional expressions spoken in a *disgusted, neutral, fearful, happy* and *sad* tone of voice than when spoken in an *angry* prosody. Likewise, listeners performed less accurate and rated themselves as less confident for utterances spoken in a *disgusted, fearful* and *sad* tone of voice than for those spoken in a *neutral* prosody. Although recognition accuracy rates were significantly higher for *neutral* than for *happy,* no significant differences in confidence ratings were observed when comparing these two emotions. *Happy* had lower recognition rates than *disgust*, yet listeners rated themselves as more confident at categorizing happy. For the other emotion comparisons, the odds of correctly detecting emotions were similar to the pattern of the differences in confidence judgements.

Odds ratio (OR) & linear contrasts (Δ) for emotion comparisons

Table S4(C) | *Odds ratio* estimates for recognition accuracy and *linear contrasts* for the pattern of the differences in the expressed confidence for the comparisons between emotion categories in semantic positive nouns

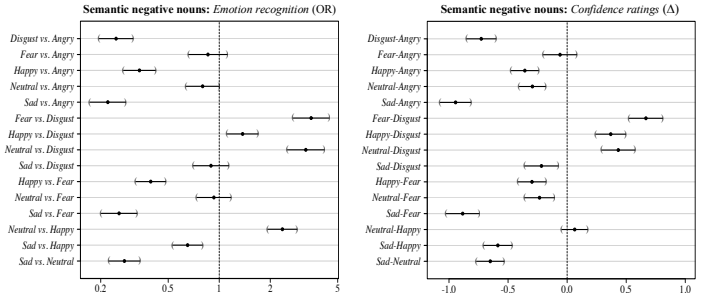
	Semantic positive nouns	
Emotion comparisons	***Emotion recognition***	***Confidence ratings***
	Odds ratio (OR)	*Difference (Δ)*
Disgust vs. Angry	0.22, [0.18; 0.27]	-0.70, [-0.81; -0.58]
Fear vs. Angry	0.46, [0.36; 0.60]	-0.20, [-0.34; -0.06]
Happy vs. Angry	0.37, [0.30; 0.46]	-0.15, [-0.26; -0.03]
Neutral vs. Angry	0.85, [0.69; 1.06]	-0.09, [-0.20; 0.03]
Sad vs. Angry	0.36, [0.29; 0.45]	-0.66, [-0.79; -0.54]
Fear vs. Disgust	2.10, [1.63; 2.70]	0.49, [0.35; 0.64]
Happy vs. Disgust	1.69, [1.38; 2.07]	0.55, [0.42; 0.67]
Neutral vs. Disgust	3.88, [3.10; 4.87]	0.61, [0.48; 0.74]
Sad vs. Disgust	1.63, [1.31; 2.03]	0.03, [-0.10; 0.16]
Happy vs. Fear	0.80, [0.64; 1.01]	0.05, [-0.08; 0.19]
Neutral vs. Fear	1.85, [1.46; 2.33]	0.11, [-0.01; 0.24]
Sad vs. Fear	0.78, [0.62; 0.98]	-0.46, [-0.60; -0.33]
Neutral vs. Happy	2.30, [1.88; 2.82]	0.06, [-0.05; 0.18]
Sad vs. Happy	0.97, [0.78; 1.19]	-0.52, [-0.65; -0.39]
Sad vs. Neutral	0.42, [0.34; 0.51]	-0.58, [-0.69; -0.46]



Note: As it can be observed, listeners were significantly less accurate and rated themselves as less confident when identifying emotional expressions spoken in a *disgusted, fearful, happy* and *sad* tone of voice than when spoken in an *angry* prosody. When comparing *neutral* to *disgust* and *sad* one can observe that both, emotion recognition accuracy and confidence ratings were significantly higher for neutral than for the other two emotional prosodies. Although recognition accuracy rates were significantly higher for *neutral* than for *fear* and *happy,* no significant differences in confidence ratings were observed when comparing these emotions. Likewise, *sad* had higher accuracy scores than *disgust*, yet there were no significant differences in confidence ratings. When comparing *sad* to *happy* the performance accuracy when categorizing these two emotions was not significantly different, however, listeners felt less confident when the positive nouns where spoken in a sad- than in a happy prosody. For the other emotion comparisons, the odds of correctly detecting emotions were similar to the pattern of the differences in confidence judgements.

Table S4(D) | *Odds ratio* estimates for recognition accuracy and *linear contrasts* for the pattern of the differences in the expressed confidence for the comparisons between emotion categories in semantic negative nouns

	Semantic negative nouns	
Emotion comparisons	***Emotion recognition***	***Confidence ratings***
	Odds ratio (OR)	*Difference (Δ)*
Disgust vs. Angry	0.25, [0.20; 0.31]	-0.73, [-0.85; -0.60]
Fear vs. Angry	0.86, [0.66; 1.11]	-0.06, [-0.20; 0.08]
Happy vs. Angry	0.34, [0.27; 0.42]	-0.36, [-0.47; -0.24]
Neutral vs. Angry	0.80, [0.64; 1.00]	-0.29, [-0.40; -0.18]
Sad vs. Angry	0.22, [0.17; 0.29]	-0.94, [-1.07; -0.81]
Fear vs. Disgust	3.48, [2.73; 4.44]	0.67, [0.53; 0.81]
Happy vs. Disgust	1.37, [1.12; 1.69]	0.37, [0.24; 0.49]
Neutral vs. Disgust	3.24, [2.53; 4.16]	0.43, [0.29; 0.57]
Sad vs. Disgust	0.89, [0.70; 1.14]	-0.22, [-0.36; -0.08]
Happy vs. Fear	0.39, [0.32; 0.48]	-0.30, [-0.41; -0.18]
Neutral vs. Fear	0.93, [0.74; 1.17]	-0.23, [-0.36; -0.11]
Sad vs. Fear	0.26, [0.20; 0.33]	-0.88, [-1.02; -0.75]
Neutral vs. Happy	2.36, [1.94; 2.87]	0.06, [-0.04; 0.17]
Sad vs. Happy	0.65, [0.53; 0.80]	-0.59, [-0.71; -0.47]
Sad vs. Neutral	0.28, [0.22; 0.34]	-0.65, [-0.77; -0.53]

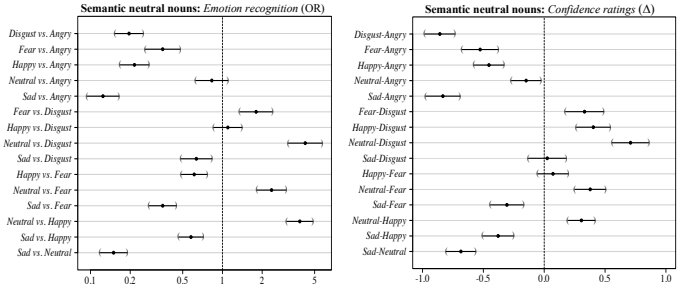


Note: As it can be observed, listeners were significantly less accurate and rated themselves as less confident when identifying emotional expressions spoken in a *disgusted, happy* and *sad* tone of voice than when spoken in an *angry* prosody. Although the recognition accuracy was not significant when comparing *fear* and *neutral* to *angry* expressions, listeners felt less confident at categorizing the stimuli spoken in a neutral- than in an angry tone of voice. Similarly, there were no significant differences in performance accuracy when comparing *sad* to *disgust* and *neutral* to *fear*, yet listeners felt more confident at categorizing the negative nouns when spoken in a disgusted- and fearful tone of voice. Despite *neutral* had significantly higher accuracy rates than *happy*, no significant differences in confidence ratings were observed. For the other emotion comparisons, the odds of correctly detecting emotions were similar to the pattern of the differences in confidence judgements.

Odds ratio (OR) & linear contrasts (Δ) for emotion comparisons

Table S4(E) | *Odds ratio* estimates for recognition accuracy and *linear contrasts* for the pattern of the differences in the expressed confidence for the comparisons between emotion categories in semantic neutral nouns

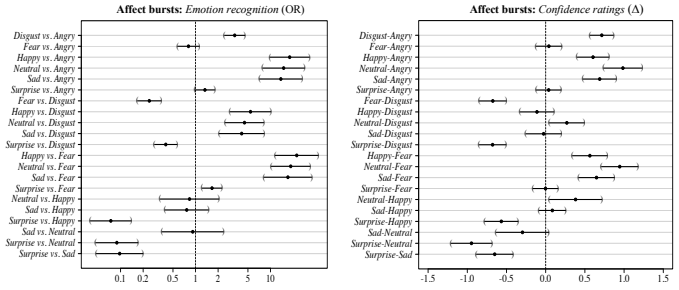
Emotion comparisons	Semantic neutral nouns	
	Emotion recognition	***Confidence ratings***
	Odds ratio (OR)	*Difference (Δ)*
Disgust vs. Angry	0.20, [0.15; 0.25]	-0.86, [-0.98; -0.74]
Fear vs. Angry	0.35, [0.26; 0.48]	-0.53, [-0.67; -0.38]
Happy vs. Angry	0.22, [0.17; 0.28]	-0.45, [-0.58; -0.33]
Neutral vs. Angry	0.83, [0.63; 1.10]	-0.15, [-0.27; -0.03]
Sad vs. Angry	0.12, [0.09; 0.16]	-0.83, [-0.97; -0.69]
Fear vs. Disgust	1.80, [1.36; 2.39]	0.33, [0.17; 0.49]
Happy vs. Disgust	1.10, [0.86; 1.41]	0.40, [0.27; 0.54]
Neutral vs. Disgust	4.24, [3.16; 5.69]	0.71, [0.56; 0.86]
Sad vs. Disgust	0.63, [0.49; 0.83]	0.02, [-0.13; 0.18]
Happy vs. Fear	0.61, [0.49; 0.76]	0.07, [-0.05; 0.20]
Neutral vs. Fear	2.36, [1.83; 3.04]	0.38, [0.25; 0.50]
Sad vs. Fear	0.35, [0.28; 0.45]	-0.31, [-0.44; -0.17]
Neutral vs. Happy	3.86, [3.08; 4.83]	0.31, [0.19; 0.42]
Sad vs. Happy	0.58, [0.47; 0.71]	-0.38, [-0.50; -0.25]
Sad vs. Neutral	0.15, [0.12; 0.19]	-0.68, [-0.80; -0.56]



Note: As it can be observed, listeners were significantly less accurate and rated themselves as less confident when identifying emotional expressions spoken in a *disgusted*, *fearful*, *happy* and *sad* tone of voice than when spoken in an *angry* prosody. Although the recognition accuracy was not significant when comparing *neutral* to *angry* expressions, listeners felt less confident at categorizing the stimuli spoken in a neutral- than in an angry tone of voice. Similarly, there were no significant differences in performance accuracy when comparing *happy* to *disgust*, yet listeners felt more confident at categorizing the neutral nouns when spoken in a happy tone of voice. Although the performance accuracy was significantly lower for *sad* than for *disgusted* prosody, as well as, for utterances spoken in a *happy* than in a *fearful* tone of voice, no significant differences in confidence ratings were observed regarding these emotion comparisons. For the other emotion comparisons, the odds of correctly detecting emotions were similar to the pattern of the differences in confidence judgements.

Table S4(F) | *Odds ratio* estimates for recognition accuracy and *linear contrasts* for the pattern of the differences in the expressed confidence for the comparisons between emotion categories in affect bursts

Emotion comparisons	Affect bursts	
	Emotion recognition	***Confidence ratings***
	Odds ratio (OR)	*Difference (Δ)*
Disgust vs. Angry	3.33, [2.44; 4.55]	0.72, [0.57;0.86]
Fear vs. Angry	0.80, [0.58; 1.12]	0.04, [-0.12; 0.20]
Happy vs. Angry	18.01, [9.86; 32.92]	0.60, [0.40; 0.80]
Neutral vs. Angry	14.97, [7.89; 28.41]	0.98, [0.74; 1.23]
Sad vs. Angry	13.73, [7.16; 26.31]	0.69, [0.48; 0.90]
Surprise vs. Angry	1.33, [0.99; 1.80]	0.04, [-0.12; 0.19]
Fear vs. Disgust	0.24, [0.17; 0.35]	-0.67, [-0.85; -0.50]
Happy vs. Disgust	5.41, [2.90; 10.10]	-0.11, [-0.33; 0.10]
Neutral vs. Disgust	4.50, [2.50; 8.10]	0.27, [0.05; 0.49]
Sad vs. Disgust	4.12, [2.07; 8.20]	-0.03, [-0.25; 0.20]
Surprise vs. Disgust	0.40, [0.28; 0.57]	-0.68, [-0.85; -0.51]
Happy vs. Fear	22.38, [11.59; 43.23]	0.56, [0.34; 0.78]
Neutral vs. Fear	18.61, [10.27; 33.71]	0.94, [0.71; 1.18]
Sad vs. Fear	17.06, [8.23; 35.35]	0.65, [0.42; 0.88]
Surprise vs. Fear	1.66, [1.23; 2.24]	-0.00, [-0.16; 0.16]
Neutral vs. Happy	0.83, [0.34; 2.04]	0.38, [0.05; 0.71]
Sad vs. Happy	0.76, [0.39; 1.48]	0.08, [-0.08; 0.25]
Surprise vs. Happy	0.07, [0.04; 0.14]	-0.57, [-0.78; -0.35]
Sad vs. Neutral	0.92, [0.36; 2.35]	-0.30, [-0.63; 0.04]
Surprise vs. Neutral	0.09, [0.05; 0.17]	-0.95, [-1.21; -0.69]
Surprise vs. Sad	0.10, [0.05; 0.20]	-0.65, [-0.88; -0.42]

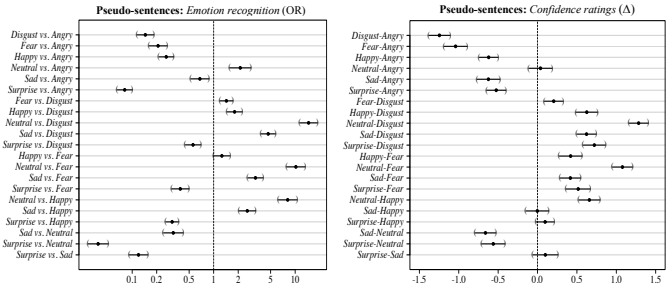


Note: As it can be observed, listeners were significantly more accurate and felt more confident at identifying emotional expressions spoken in a *disgusted*, *happy*, *neutral* and *sad* tone of voice than when spoken in an *angry* prosody. While performance accuracy was significantly higher for *happy* and *sad* than for *disgust*, no significant differences in confidence ratings were observed when comparing these emotions. Utterances spoken in a *surprised* tone of voice had significantly lower recognition rates than when spoken in a *fearful* prosody, yet, no significant differences in confidence ratings were observed. When comparing *neutral* to *happy*, the recognition rates were not significant, however, listeners felt more confident at categorizing neutral than happy expressions. For the other emotion comparisons, the odds of correctly detecting emotions were similar to the pattern of the differences in confidence judgements.

Odds ratio (OR) & linear contrasts (Δ) for emotion comparisons

Table S4(G) | *Odds ratio* estimates for recognition accuracy and *linear contrasts* for the pattern of the differences in the expressed confidence for the comparisons between emotion categories in pseudo-sentences

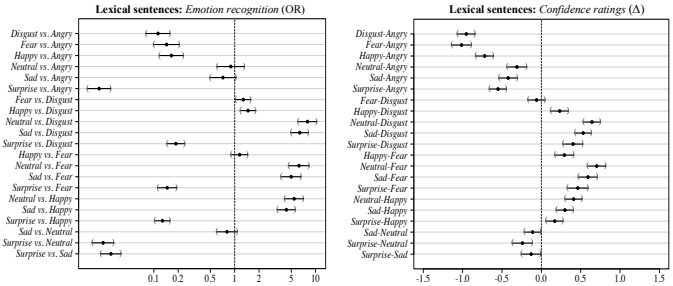
	Pseudo-sentences	
Emotion comparisons	***Emotion recognition***	***Confidence ratings***
	Odds ratio (OR)	*Difference (Δ)*
Disgust vs. Angry	0.15, [0.11; 0.18]	-1.25, [-1.38; -1.11]
Fear vs. Angry	0.21, [0.16; 0.27]	-1.04, [-1.19; -0.89]
Happy vs. Angry	0.26, [0.21; 0.32]	-0.62, [-0.74; -0.50]
Neutral vs. Angry	2.12, [1.57; 2.86]	0.04, [-0.11; 0.18]
Sad vs. Angry	0.68, [0.52; 0.87]	-0.62, [-0.77; -0.48]
Surprise vs. Angry	0.08, [0.07; 0.10]	-0.52, [-0.64; -0.40]
Fear vs. Disgust	1.43, [1.19; 1.72]	0.21, [0.09; 0.32]
Happy vs. Disgust	1.80, [1.46; 2.23]	0.62, [0.49; 0.76]
Neutral vs. Disgust	14.61, [11.38; 18.75]	1.28, [1.16; 1.40]
Sad vs. Disgust	4.67, [3.81; 5.72]	0.62, [0.50; 0.74]
Surprise vs. Disgust	0.56, [0.45; 0.69]	0.72, [0.58; 0.86]
Happy vs. Fear	1.26, [1.00; 1.59]	0.42, [0.28; 0.56]
Neutral vs. Fear	10.19, [7.86; 13.23]	1.08, [0.95; 1.20]
Sad vs. Fear	3.26, [2.63; 4.03]	0.42, [0.29; 0.55]
Surprise vs. Fear	0.39, [0.31; 0.50]	0.52, [0.36; 0.67]
Neutral vs. Happy	8.09, [6.22; 10.53]	0.66, [0.53; 0.79]
Sad vs. Happy	2.58, [2.04; 3.27]	-0.00, [-0.15; 0.14]
Surprise vs. Happy	0.31, [0.26; 0.37]	0.10, [-0.02; 0.21]
Sad vs. Neutral	0.32, [0.24; 0.42]	-0.66, [-0.79; -0.53]
Surprise vs. Neutral	0.04, [0.03; 0.05]	-0.56, [-0.71; -0.41]
Surprise vs. Sad	0.12, [0.09; 0.15]	0.10, [-0.06; 0.26]



Note: As it can be observed, listeners were significantly less accurate and rated themselves as less confident when identifying emotional expressions spoken in a *disgusted, fearful, happy*, *sad* and *surprised* tone of voice than when spoken in an *angry* prosody. When comparing *neutral* to *angry*, the recognition rates were significantly higher for utterances spoken in a neutral tone of voice, yet, there were no significant differences in confidence ratings when comparing these two emotions. Similarly, listeners were more accurate at categorizing utterances spoken in a *happy*- than in a *sad* prosody or in a *sad* than in a *surprised* tone of voice, yet again, there were no significant differences in confidence ratings when comparing these emotions. Although listeners were less accurate to categorize *surprise* than *disgust* and *fear*, they rated themselves as more confident at categorizing surprise. For the other emotion comparisons, the odds of correctly detecting emotions were similar to the pattern of the differences in confidence judgements.

Table S4(H) | *Odds ratio* estimates for recognition accuracy and *linear contrasts* for the pattern of the differences in the expressed confidence for the comparisons between emotion categories in lexical sentences

	Lexical sentences	
Emotion comparisons	***Emotion recognition***	***Confidence ratings***
	Odds ratio (OR)	*Difference (Δ)*
Disgust vs. Angry	0.11, [0.08; 0.16]	-0.95, [-1.06; -0.84]
Fear vs. Angry	0.14, [0.10; 0.20]	-1.01, [-1.13; -0.89]
Happy vs. Angry	0.16, [0.12; 0.23]	-0.72, [-0.83; -0.61]
Neutral vs. Angry	0.89, [0.61; 1.31]	-0.31, [-0.43; -0.19]
Sad vs. Angry	0.72, [0.50; 1.03]	-0.42, [-0.53; -0.31]
Surprise vs. Angry	0.02, [0.02; 0.03]	-0.55, [-0.65; -0.45]
Fear vs. Disgust	1.28, [1.04; 1.57]	-0.06, [-0.16; 0.04]
Happy vs. Disgust	1.46, [1.19; 1.80]	0.23, [0.13; 0.34]
Neutral vs. Disgust	7.94, [6.14; 10.27]	0.64, [0.54; 0.75]
Sad vs. Disgust	6.36, [5.03; 8.06]	0.53, [0.43; 0.63]
Surprise vs. Disgust	0.18, [0.15; 0.24]	0.40, [0.28; 0.52]
Happy vs. Fear	1.15, [0.91; 1.44]	0.29, [0.18; 0.41]
Neutral vs. Fear	6.22, [4.72; 8.21]	0.70, [0.59; 0.81]
Sad vs. Fear	4.99, [3.79; 6.57]	0.59, [0.48; 0.71]
Surprise vs. Fear	0.15, [0.11; 0.19]	0.46, [0.33; 0.59]
Neutral vs. Happy	5.43, [4.20; 7.02]	0.41, [0.31; 0.51]
Sad vs. Happy	4.35, [3.40; 5.57]	0.30, [0.20; 0.40]
Surprise vs. Happy	0.13, [0.10; 0.16]	0.17, [0.07; 0.27]
Sad vs. Neutral	0.80, [0.60; 1.07]	-0.11, [-0.21; -0.01]
Surprise vs. Neutral	0.02, [0.02; 0.03]	-0.24, [-0.36; -0.12]
Surprise vs. Sad	0.03, [0.02; 0.04]	-0.13, [-0.25; -0.01]

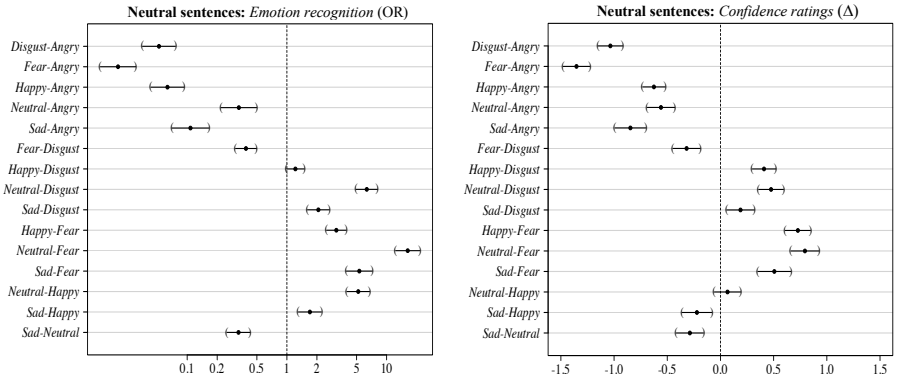


Note: As it can be observed, listeners were significantly less accurate and rated themselves as less confident when identifying emotional expressions spoken in a *disgusted, fearful, happy* and *surprised* tone of voice than when spoken in an *angry* prosody. Although the recognition accuracy was not significant when comparing *neutral* and *sad* to angry expressions, listeners felt less confident when categorizing the stimuli spoken in a neutral- and sad- than in an angry tone of voice. Listeners were less accurate to categorize *surprise* than *disgust*, *fear* and *happy*, yet, they rated themselves as more confident at categorizing surprise. No significant differences in emotion recognition accuracy were observed when comparing *sad* to *neutral*, however, listeners felt less confident when categorizing sad- than neutral prosody. For the other emotion comparisons, the odds of correctly detecting emotions were similar to the pattern of the differences in confidence judgements.

Odds ratio (OR) & linear contrasts (Δ) for emotion comparisons

Table S4(I) | *Odds ratio* estimates for recognition accuracy and *linear contrasts* for the pattern of the differences in the expressed confidence for the comparisons between emotion categories in neutral sentences

Emotion comparisons	**Neutral sentences**	
	Emotion recognition	***Confidence ratings***
	Odds ratio (OR)	*Difference (Δ)*
Disgust vs. Angry	0.05, [0.04; 0.08]	-1.04, [-1.15; -0.92]
Fear vs. Angry	0.02, [0.01; 0.03]	-1.35, [-1.48; -1.23]
Happy vs. Angry	0.06, [0.04; 0.09]	-0.63, [-0.73; -0.52]
Neutral vs. Angry	0.33, [0.22; 0.50]	-0.56, [-0.69; -0.43]
Sad vs. Angry	0.11, [0.07; 0.16]	-0.85, [-0.99; -0.70]
Fear vs. Disgust	0.39, [0.31; 0.49]	-0.32, [-0.45; -0.19]
Happy vs. Disgust	1.22, [0.99; 1.49]	0.41, [0.30; 0.52]
Neutral vs. Disgust	6.32, [4.95; 8.08]	0.48, [0.36; 0.59]
Sad vs. Disgust	2.06, [1.61; 2.66]	0.19, [0.06; 0.32]
Happy vs. Fear	3.13, [2.50; 3.92]	0.73, [0.61; 0.85]
Neutral vs. Fear	16.27, [12.31; 21.51]	0.79, [0.66; 0.93]
Sad vs. Fear	5.32, [3.95; 7.17]	0.51, [0.35; 0.66]
Neutral vs. Happy	5.20, [4.00; 6.75]	0.07, [-0.06; 0.19]
Sad vs. Happy	1.70, [1.30; 2.22]	-0.22, [-0.36; -0.08]
Sad vs. Neutral	0.33, [0.25; 0.43]	-0.29, [-0.41; -0.16]



Note: As it can be observed, listeners were significantly less accurate and rated themselves as less confident when identifying emotional expressions spoken in a *disgusted, fearful, happy*, *neutral* and *sad* tone of voice than when spoken in an *angry* prosody. Although the recognition accuracy was significantly higher for *neutral* than for *happy* expressions, there were no significant differences in confidence ratings when comparing these two emotions. The recognition rates were significantly higher for utterances spoken in a *sad* than in a *happy* tone of voice, however, listeners felt more confident at categorizing happy than sad expressions. For the other emotion comparisons, the odds of correctly detecting emotions were similar to the pattern of the differences in confidence judgements.

Chapter 4

Hormonal and Modality Specific Effects on Males' Emotion Recognition Ability

Abstract

Successful emotion recognition is a key component of our socio-emotional communication skills. However, little is known about the factors impacting males' accuracy in emotion recognition tasks. This pre-registered study examined potential candidates, focusing on the modality of stimulus presentation, emotion category, and individual hormone levels. We obtained accuracy and reaction time scores from 312 males who categorized voice, face and voice-face stimuli for nonverbal emotional content. Results showed that recognition accuracy was significantly higher in the audio-visual than in the auditory or visual modality. While no significant association was found for testosterone and cortisol alone, the effect of the interaction with recognition accuracy and reaction time was significant, but small. Our results establish that audio-visual congruent stimuli enhance recognition accuracy and provide novel empirical support by showing that the interaction of testosterone and cortisol modulate to some extent males' accuracy and response times in emotion recognition tasks.[1]

Keywords: Emotion Recognition, Prosody, Facial Expressions, Testosterone, Cortisol, Dual-hormone hypothesis

[1]Lausen, A., Broering, C., Penke, L., & Schacht, A. (submitted). Hormonal and modality specific effects on males' emotion recognition ability.

4.1 Introduction

Emotion recognition is a basic skill thought to carry clear advantages for predicting behavior, as well as forming and maintaining social bonds (Soto & Levenson, 2009). Intriguingly, research on sex differences highlights that males are less accurate than females when completing emotion recognition tasks (e.g., Hall, 1984; A. E. Thompson & Voyer, 2014). However, effect sizes are comparably small and multiple factors known to impact the ability to recognize emotions have yet to be fully controlled for (see, Chaplin, 2015; A. H. Fischer & LaFrance, 2015; Hall et al., 2000; Hyde, 2014; Schirmer, 2013, for an overview regarding explanations for sex-based behavior patterns). The ability to correctly interpret emotional expressions forms the basis of social interactions and personal relationships (e.g., A. H. Fischer & Manstead, 2008; Keltner & Kring, 1998), yet there is a lack of direct evidence for reasons why males have an often assumed disadvantage when it comes to accurately recognizing emotions. Therefore, the main aim of this study was to systematically investigate potential factors that might impact males' ability to recognize emotions.

One of the factors supposed to impact emotion recognition is the modality of stimulus presentation (Hall, 1984). In many everyday situations, judgments about others' emotional states require the integration of information from various sensory modalities making use of different cues such as facial expressions, tone of voice (i.e., prosody), or body language (Klasen et al., 2014). Thus, it has been argued that emotion recognition is a multimodal event (Piwek, Pollick, & Petrini, 2015). Indeed, a growing number of studies have pointed out that in emotion recognition tasks the stimuli presented in isolation (i.e., visual or auditory) have lower accuracy scores and slower response times than the audiovisual presentation of emotional expressions (e.g., Bänziger et al., 2009; Collignon et al., 2008; De Gelder & Vroomen, 2000; Jessen, Obleser, & Kotz, 2012; Kreifelts et al., 2007; Paulmann & Pell, 2011). Research on unimodal emotion recognition reports that emotions are better recognized from faces than from voices (e.g., Waaramaa, 2017). However, these observations are often contradictory (e.g., Kraus, 2017). Furthermore, previous research in the unimodal domains highlighted that specific emotions are not recognized equally well in the auditory and visual modality. In studies on the vocal channel, participants were faster and most accurate to recognize anger (e.g., Chronaki et al., 2018; Cornew et al., 2010; Juslin & Laukka, 2003), while in studies on facial expressions, happiness was shown to be recognized more accurately and faster than any other emotion (e.g., Elfenbein & Ambady, 2002a; Kosonogov & Titova, 2018; Montagne, Kessels, De Haan, & Perrett, 2007; Nummenmaa & Calvo, 2015; Palermo & Coltheart, 2004; Wells et al., 2016; L. M. Williams et al., 2009). Despite these converging patterns, it is as yet not possible to make definite claims regarding the advantage of certain emotional categories because, at least within the vocal domain, recognition accuracy (RA) was found to be strongly influenced by the type of stimulus used (see, Lausen, Hammerschmidt, & Schacht, 2019, for an overview). Whether the voice is a more reliable source than the face in emotion recognition tasks has been rarely pursued, and results are limited to specific emotions, paradigms, as well as, by a number of methodological differences between studies. Thus, until further evidence

regarding RA within specific sensory modalities and emotional categories is provided, the direction of these effects remains an open question.

A recently emphasized influence on the ability to recognize emotions concerns potential effects of steroid hormones, such as testosterone (Gignell et al., 2019). Testosterone (T) receptors are distributed throughout the nervous system with high concentrations in areas associated with emotional processing such as the hypothalamus and amygdala (see Gignell et al., 2019, for details). However, only few studies have assessed the influence of T concentrations on emotion recognition in both sexes and an even smaller subsection has specifically addressed the impact of T levels on males' ability to recognize emotions. For example, an fMRI study by Derntl et al. (2009) investigated the influence of blood T levels on males' RA in an explicit emotion recognition task. Results showed increased amygdala activity in individuals with high T levels during the presentation of fearful and angry faces. In addition, the authors found that reaction times (RTs) to fearful male faces negatively correlated with T level concentrations. However, no correlation was found between RA and T levels. Subsequent studies reported a negative correlation between salivary T levels and emotion recognition in male adolescent groups (Fujisawa & Shinohara, 2011) or found a positive correlation between higher levels of T and emotion recognition (Vongas & Al Hajj, 2017). By presenting participants with emotional facial expressions at two different intensity levels (i.e., 50% and 100%), Rukavina et al. (2018) found that RA decreases when salivary T is high, especially for full-blown expressions of sadness and for disgust when presented at 50% intensity. Based on these findings, the authors concluded that RA decreases with increasing levels of T.

These contradictory findings are likely the result of a number of methodological differences such as insufficient statistical power (i.e., sample sizes ranging from 21 to 84 males), T assessment from blood or saliva, as well as storage and analyses of hormone samples (see, Schultheiss et al., 2019, for details).Another possible explanation for the discrepancies is that another hormone, cortisol (C), may constrain T influence on emotion recognition. C, an end product of the hypothalamic-pituitary-adrenal (HPA) axis, was found to inhibit T by reducing hypothalamic-pituitary-gonadal (HPG) activity and blocking androgen receptors (see, Sarkar, Mehta, & Josephs, 2019; Viau, 2002, for details). To reconcile mixed findings on the roles of T and C in human social behavior, Mehta and Josephs (2010) proposed the *dual-hormone hypothesis*. According to this hypothesis T predicts a wide range of behaviors, but only under the condition that C concentrations are low. If C concentrations are high, the T-behavior association is supposed to be attenuated (Carré & Mehta, 2011; Mehta & Prasad, 2015). This hypothesis was supported in a variety of studies, which demonstrated that across different psychological domains the interaction between T and C influences empathy, as well as, dominant, status-relevant, risk-taking and antisocial behavior (see Sarkar et al., 2019, for an overview). However, it should be noted that other studies report only small effects (Dekkers, 2018), null-findings (e.g., Mazur & Booth, 2014), and even reversed patterns (i.e., T was related to status-relevant behavior or facial dominance for high but not low C; Kordsmeyer et al., 2019; Welker, Lozoya, Campbell, Neumann, & Carré, 2014) for the dual-hormone hypothesis. Considering the interaction between the

HPG and HPA axes might nevertheless lead to more reliable predictions regarding emotion recognition than the assumption of a single-hormone association (Carré & Mehta, 2011; Sarkar et al., 2019).

Based on the above-mentioned findings, the present study had three major aims. Firstly, it aimed at examining whether males' RA is influenced by the modality of stimulus presentation. We hypothesized that RA would be better in the audio-visual modality than in the auditory or visual modality (1a), and lower in the visual compared to the auditory modality (1b). Second, we aimed to replicate previous findings by examining the extent RA and RTs vary across discrete emotion categories as a function of modality (e.g., Lambrecht et al., 2014). Specifically, we expected higher accuracy scores and faster RTs for disgusted, fearful and sad expressions in the audio-visual than in both the auditory and the visual modality (2a). We also hypothesized that angry expressions would be identified faster and with higher accuracy in the vocal compared to the facial domain, while for happy expressions we expected the reverse pattern (2b). A third aim was to alleviate some of the methodological flaws of previous research by using a large sample size to examine whether variations in males' ability to recognize emotions are due to T level concentrations. We expected a negative correlation between T and RA (3a), and that participants with high levels of T would specifically react faster to angry and fearful expressions (3b)[2]. In addition, we conducted an exploratory analysis on the associations between C and RA, C and RT, as well as on the relationship between RA or RT and the interaction between T and C levels.

4.2 Method

The study was approved by the ethics committee of the Georg-Elias-Mueller-Institute of Psychology (University of Goettingen), and conducted in accordance with the ethical principles formulated in the Declaration of Helsinki (2013). Participants gave informed consent and were reimbursed with course credit or 8 Euros per hour.

4.2.1 Participants

A total of 312 males (age range 18-36 years; $M_{\text{Age}} = 24.3$, $SD = 3.7$) were recruited on the university campus using flyers and the Institute of Psychology participant database (*ORSEE*, www.orsee.org), as well as, by posts on the social media site *Facebook* and the online platform *Stellenwerk Jobportal* University Goettingen (www.stellenwerk-goettingen.de). Of the 312 recruited subjects, 30 participants were excluded from analysis due to self-reported hearing problems, psychiatric or neurological disorders, or intake of psychotropic/ hormone medication. After these exclusions, a total of 282 participants with a mean age of 24.3 years ($SD = 3.8$) were included in the analysis.

[2] All hypotheses tested in the current paper have been pre-registered (osf.io/w2tgr). This pre-registration contained further hypotheses that are not part of the present paper.

4.2.2 Stimulus material

Stimuli were displayed under three experimental modality conditions: auditory, visual and audio-visual. In each experimental condition, stimuli were presented in one of the emotions of interest (i.e., anger, disgust, fear, happiness, sadness) as well as in a neutral state (i.e., baseline expression).

Audio stimuli

The audio stimuli consisted of pseudo-speech (i.e., pseudo-words, pseudo-sentences) and non-verbal vocalizations (i.e., affect bursts). We decided to use pseudo-speech (i.e., a language devoid of meaning) and non-verbal vocalizations as they have been argued to capture the pure effects of emotional prosody independent of lexical-semantic cues and, to be an ideal tool when investigating the expression of emotional information when there is no concurrent verbal information present (Banse & Scherer, 1996; Pell et al., 2015). The stimuli were sampled from well-established databases or provided by researchers who developed their own stimulus materials. We validated all stimuli in a previous study (cf. Lausen et al., 2019; Lausen & Schacht, 2018) and selected only a subset of stimuli (i.e., with the highest accuracy) from each database (see ***Table 1*****).** The physical volume of

Table 1 | Description of audio materials

Database	Speakers	Emotions	Nature of material	Number of stimuli selected	Total stimuli
Magdeburg Prosody Corpus (Wendt & Scheich, 2002)	2 actors (1 male/1female)	Anger, disgust, fear, happiness, sadness, neutral	Pseudo-words	4	48
Paulmann Prosodic Stimuli (Paulmann & Kotz, 2008; Paulmann et al., 2008)	2 actors (1 male/1female)		Pseudo-sentences	4	48
Montreal Affective Voices (Belin et al., 2008)	8 actors (4 male/4female)		Affect bursts		48

stimulus presentations across the nine laptops used in the experiment was controlled by measuring sound volume of the practice trials with a professional sound level meter, Nor140 (Norsonic, 2010, Lierskogen, Norway). No significant difference in volume intensity was observed [$F_{(8,40)} = 1.546$, $p = 0.173$].

Visual stimuli

Visual stimuli consisted of 24 frontal face photographs (12 males/12 females) extracted from the *Radboud Faces Database* (Langner et al., 2010). The presentation time of the faces was matched to the length of the voice stimuli (i.e., from 319 ms to 4821 ms). A gray ellipsoid mask, ensuring a uniform figure/ground contrast surrounded the stimuli, with only the internal area of the face visible (9x14 cm, width and height). The stimuli were presented in colour and corrected for luminance across emotion conditions [$F_{(5,137)} = 0.200$, $p = 0.962$], using *Adobe Photoshop CS6* (Version 13.0.1, 2012, San Jose, CA).

Audio-visual stimuli

The voice stimuli were simultaneously presented with the face stimuli. Using *Adobe Premiere Pro CS6* (Version 6.0.5) videos were created, matching face and voice stimuli for sex and emotion category.

4.2.3 Procedure, experimental task and saliva samples

Participants were informed that the study required them to provide two saliva samples over a period of about two hours. A day before the main experiment, they were sent an email instructing them to abstain from sports and the consumption of alcohol, drugs or unnecessary medication on the day of the study. Furthermore, they were instructed not to consume drinks containing caffeine within three hours of the experiment and to refrain from eating, drinking (except water), smoking and brushing their teeth within one hour of the experiment. Adherence to these instructions was assessed using a screening questionnaire (Schultheiss & Stanton, 2009). As individual differences in peak hormone levels measured in the morning have been argued to be a better predictor of behavioural responses to emotional stimuli than measurements later in the day (Schultheiss & Stanton, 2009), the designated time slot for testing was between 9:00am to 11:00am.

Participants were tested in groups of up to nine individuals. On the day of the study, after completing the consent form, participants received oral and written instructions about the procedure of the experiment and the collection of saliva samples. The saliva samples were collected before (T1) and after (T2) the *Emotion Recognition Task*[3]. The experiment was programmed using *Python* (Version 2.7.0, Python Software Foundation, Beaverton, OR) and run on a *Dell Latitude E5530* Laptop with a 15.6 LCD display screen. The audio stimuli were presented binaurally via headphones (*Bayerdynamic DT 770 PRO*).

Emotion recognition task

The emotion recognition task consisted of three blocks, each block displaying one of the three experimental conditions: auditory, visual, and audio-visual. Each experimental condition contained 144 stimuli. A permutation was applied to randomize the order in which the experimental conditions were presented to the participants. Six different permutations were created, and each permutation was allocated randomly in blocks of six participants. The order of the stimuli within each experimental condition was completely randomized. The audio and visual stimuli were matched for duration, sex, and emotion category (see *Table S1* in supplementary material for an example of how the audio and visual stimuli were matched). Before each experimental condition, participants were familiarized with the task in a short training session comprised of three stimuli. Each trial began with a blank screen followed by a fixation cross. Following the presentation of a stimulus, a circular answer display appeared, containing all six categories of interest (i.e., anger, disgust, fear, happiness, sadness, neutral) and the selection cursor, which appeared in the centre of the display. The sequence of the emotion labels was randomized for each participant and

[3]The data reported in this paper was obtained within the confines of a larger study. The experiment began with a short demographic questionnaire followed by the *Screening Questionnaire* (Schultheiss & Stanton, 2009), *Multi-Motive Grid* (MMG, Sokolowski, Schmalt, Langens, & Puca, 2000) and *Positive and Negative Affect Schedule* (PANAS Breyer & Bluemke, 2016). Next, the first saliva sample (T1) was taken. After a short break, the **Emotion Recognition Task** ensued, followed by PANAS, and the collection of the second saliva sample (T2). The saliva samples were collected approximately 10 minutes before and after the emotion recognition task. The experiment ended with the completion of *Multifaceted Empathy Test* short-form (MET Dziobek et al., 2008) and *Big Five Inventory* (BFI, Danner et al., 2016). As MMG, PANAS, MET and BFI are not relevant to the present manuscript they are not further reported.

remained the same throughout the task. Participants had to select an emotion category, using the mouse to move the cursor, before the next stimulus was presented. Reaction times were measured, starting with the onset of the answer display and ending with the participant's response. ***Figure 1*** displays the time course of the emotion recognition task.

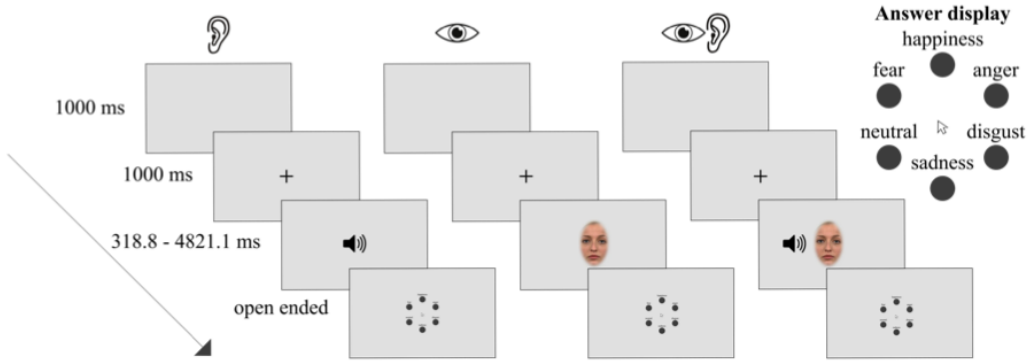


Figure 1 | Emotion recognition task
Each trial began with a blank screen (shown for 1000ms) which was followed by a fixation-cross appearing at the center of the screen (for 1000ms) at which participants were asked to fixate throughout the trial. The presentation of the stimuli was initiated by pressing the *Spacebar*-key at the beginning of each block. After the presentation of the stimulus a circular answer display containing all six categories of interest (i.e., anger, disgust, fear, happiness, neutral and sadness) and the selection cursor (which appeared in the center of the display) were presented. The responses were made by using the mouse to move the cursor. Reaction times were measured, starting with the onset of the answer display and ending with the participant's response. There was no time limit for emotion judgments. Participants could hear/see the stimulus only once. At the end of each block a visual message in the center of the screen instructed participants to take a break if they wished to or to press the *Spacebar*-key to proceed with the next block.

Saliva samples

The two saliva samples (2 ml per sample) were collected from each participant via passive drool through a straw (Schultheiss, Schiepe, & Rawolle, 2012) into an IBL SaliCap sampling device. These plastic vials were stored frozen at -80°C until shipment on dry ice to the Endocrinology Laboratory at Technical University of Dresden. At this facility, the samples were analyzed for T and C levels via chemiluminescence immunoassays with high sensitivity (IBL International, Hamburg, Germany). The intra- and inter-assay coefficients of variation for T were $< 11\%$ and for C $< 8\%$. For T the variance between participants was 14.81% and 3.85% within participants with an intra-class correlation coefficient (ICC) of 79.35%, while for C the variance between participants was 23.78% and 28.20% within participants with an ICC of 45.74%. As the distributions of T and C were positively skewed ($T_{skewness} = 1.56$; $C_{skewness} = 1.49$) a log-transformation was performed (e.g., Mehta et al. 2015). The log-transformation reduced skewness substantially [log(T) skewness = -0.06; log(C) skewness = 0.01]. Outliers were winsorized to $\pm$ 3 standard deviations (Mehta et al. 2015).

4.2.4 Study design and power analysis

A balanced within-subjects factorial design was fitted to assess males' judgments of emotions. The design was balanced for modalities, emotion categories and encoder sex in each stimulus type. Independent within-participant factors were *modalities*, *emotion categories*, *stimuli types* and *encoder sex*. Independent between-participant variables were T and C. Dependent variables were RA and RT.

A target sample size of 231 males was determined using an approximate correlation power analysis, Bonferroni-corrected for multiple testing ($r = .25$; $\alpha = .05/20$; $1 - \beta = .80$). To account for possible attrition, the sample size was increased by a minimum of 14%.

4.2.5 Statistical analysis

In line with our preregistration, the primary analysis for our first and second hypotheses was performed using *Friedman-* and *Wilcoxon-rank-sum* tests. For the correlation between the dependent variables (RA, RT) and T levels we ran *Spearman* correlations (H3a, b).

The exploratory analyses of the quantitative variables T and C were performed using *generalized linear models* (*quasi-binomial logistic regression*) for the binary response variable emotion recognition and linear models for the response variable reaction time, which was normalized by log transformation. To obtain a more reliable value and to cover the observation interval, the two baseline measures for T and C were averaged (Kordsmeyer et al., 2019). The dispersion parameter of the quasi-binomial model accounted for dependencies caused by repeated measurements within the participants. Modality and emotion category were fitted as nominal variables and stimulus duration as quantitative variable. The interaction of the quantitative variables T and C was fitted by the product of both variables as an additional predictor. *Tertiles* for both variables, T and C, were fitted to investigate more general interaction patterns and to reduce the influence of T and C extreme values on the model equation. Chi-square tests of the deviance analysis and F-tests of the analysis of variance were used to analyse effects of predictor variables. In the quasi-binomial logistic regression, odds ratio (OR) were used to compare emotion recognition accuracies. RTs were compared by the difference of the means. Tukey's method of multiple pairwise comparisons was used to compute simultaneous 95% confidence intervals for both, OR and mean differences.

For the descriptive analysis of the data, *relative frequencies*, *confusion matrices* and *Wagner's* (1993) *unbiased hit rate* (H_u), which is the rate of correctly identified stimuli multiplied by the rate of correct judgments of the stimuli, were calculated. The data was analyzed using the R language and environment for statistical computing and graphics version 3.4.3 (R Core Team, 2017) and the integrated environment R-Studio version 1.0.153 (used packages: *pwr*; *MASS*; *coin*; *glm*; *multcomp*; *mvtnorm*; *ggplot2*).

4.3 Results

4.3.1 Descriptive analysis

Audio-visual emotional expressions were recognized with approximately 90% accuracy (lowest identification rate 89% for disgust). Angry expressions were recognized with better accuracy from the voice (90%) than the face (82%). Conversely, for fearful, happy and sad expressions accuracy scores were higher when presented visually (85% $\leq$ *accuracy scores* $\leq$ 99%) than auditorily (72% $\leq$ *accuracy scores* $\leq$ 77%). Neutral expressions had high accuracy scores in all three conditions of stimulus presentation (90% $\leq$ *accuracy scores*

$\leq$ 95%). Participants were faster at categorizing disgust, fear, happy, sad and neutral expressions in the visual and audio-visual modalities [median (*Md*) values between 1.03 sec. to 1.46 sec.] than in the auditory modality [*Md* values between 1.50 sec. to 1.95 sec.]. Although the RTs for disgusted, sad and neutral expressions were similar in the visual and audio-visual modalities, participants were slightly faster at categorizing fear and happy in the visual- than audio-visual modality. For angry expressions, the RTs were much shorter in the audio-visual (1.23 sec.) than in the auditory and visual modality, but much longer in the visual- (1.53 sec.) than in the auditory modality (1.47 sec.). ***Figure 2*** illustrates participants RA (panel A) and RT (panel B) by modality and emotion categories.

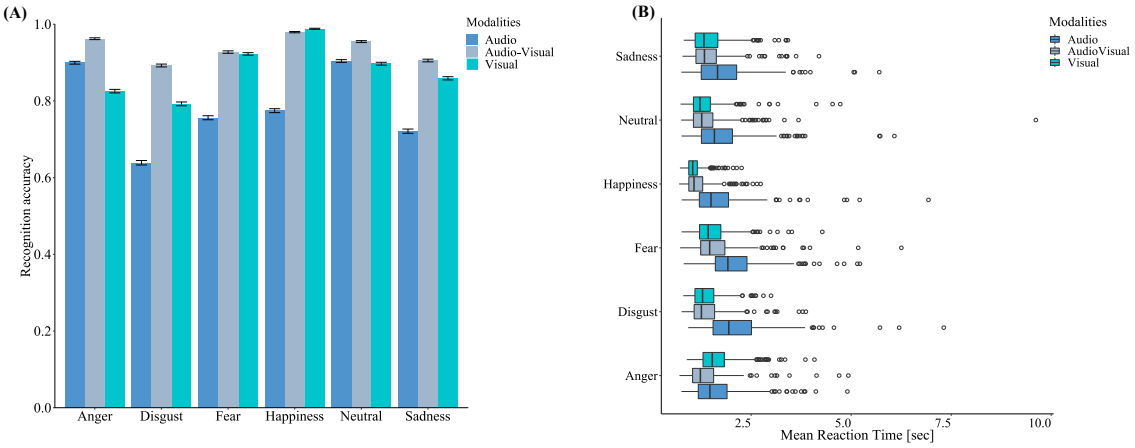


Figure 2 | Recognition accuracy (RA) and reaction times (RTs) by modality and emotion categories
The bar charts **(panel A)** display RA, while the boxplots **(panel B)** illustrate the mean RT distributions. Error bars represent the standard error. The boxplots indicate that the distributions of RT are right skewed.

In all three modalities participants often misclassified happy and sad expressions as neutral. In the auditory and audio-visual modalities angry was mistaken for fearful, neutral for angry and fearful for sad. In the visual modality fear was confused with disgust, whereas anger and neutral were confused with sadness. Participants frequently misclassified disgust with anger in the visual and audio-visual modalities, while in the auditory modality disgust was mistaken for neutral. The error classification patterns along with the unbiased hit rates are presented in ***Table 2***.

4.3.2 Main analysis

Performance accuracy in the three modalities [Aim 1]

Participants' RA was significantly influenced by the modality of stimulus presentation (*Friedman test*: $\chi^2_{(2)} = 448.56$, $p < 0.001$). The results of *Wilcoxon-rank-sum* test indicated that RA was significantly higher in the audio-visual modality than in the visual- ($z = 12.99$, $p < 0.001$, ${}_{95\%}CI = [0.052; 0.062]$, effect size ($r$) = 0.774) or auditory modality ($z = 14.525$, $p < 0.001$, ${}_{95\%}CI = [0.146; 0.163]$, $r = 0.865$). Participants' were also significantly more accurate at discriminating emotions when making judgments on visual- than on audio stimuli ($z = 13.553$, $p < 0.001$, ${}_{95\%}CI = [0.090; 0.108]$, $r = 0.807$). ***Figure 3*** illustrates RA in the three conditions of stimulus presentation.

Table 2 | Confusion Matrices and unbiased hit rates (H_u) for participants judgments of emotion categories

Modality	Emotions portrayed	Emotion judgments							
		Anger	*Disgust*	*Fear*	*Happiness*	*Neutral*	*Sadness*	*Total*	H_u
Auditory	*Anger*	**6089**	59	267	152	175	26	6768	.766
	Disgust	347	**4324**	438	280	815	564	6768	.590
	Fear	162	173	**5118**	96	406	813	6768	.621
	Happiness	116	27	15	**5243**	1335	32	6768	.665
	Neutral	339	52	62	159	**6119**	37	6768	.549
	Sadness	97	50	335	175	1230	**4881**	6768	.554
	Total	7150	4685	6235	6105	10080	6353	40608	—
Visual	*Anger*	**5587**	244	194	6	234	503	6768	.638
	Disgust	1288	**5363**	48	13	41	15	6768	.704
	Fear	51	282	**6245**	14	73	103	6768	.847
	Happiness	6	2	11	**6689**	59	1	6768	.967
	Neutral	167	15	47	102	**6071**	365	6767*	.791
	Sadness	135	134	262	11	412	**5814**	6768	.734
	Total	7234	6040	6807	6835	6890	6801	40607	—
Audio-visual	*Anger*	**6513**	46	91	8	71	39	6768	.860
	Disgust	505	**6040**	69	14	81	59	6768	.858
	Fear	39	155	**6277**	9	92	196	6768	.873
	Happiness	5	2	7	**6629**	121	4	6768	.969
	Neutral	170	11	25	35	**6462**	65	6768	.859
	Sadness	55	27	196	9	353	**6128**	6768	.855
	Total	7287	6281	6665	6704	7180	6491	40608	—
Across all 3 modalities	*Anger*	**18189**	349	552	166	480	568	20304	.752
	Disgust	2140	**15727**	555	307	937	638	20304	.716
	Fear	252	610	**17640**	119	571	1112	20304	.780
	Happiness	127	31	33	**18561**	1515	37	20304	.864
	Neutral	676	78	134	296	**18652**	467	20303*	.710
	Sadness	287	211	793	195	1995	**16823**	20304	.709
	Total	21671	17006	19707	19644	24150	19645	121823	—

Note: Frequencies of correctly judged portrayals are given on the main diagonal in boldface type. *If the number is less than the planned number of emotion judgments that is due to recording failure. H_u = the rate of correctly identified stimuli multiplied by the rate of correct judgments of the stimuli.

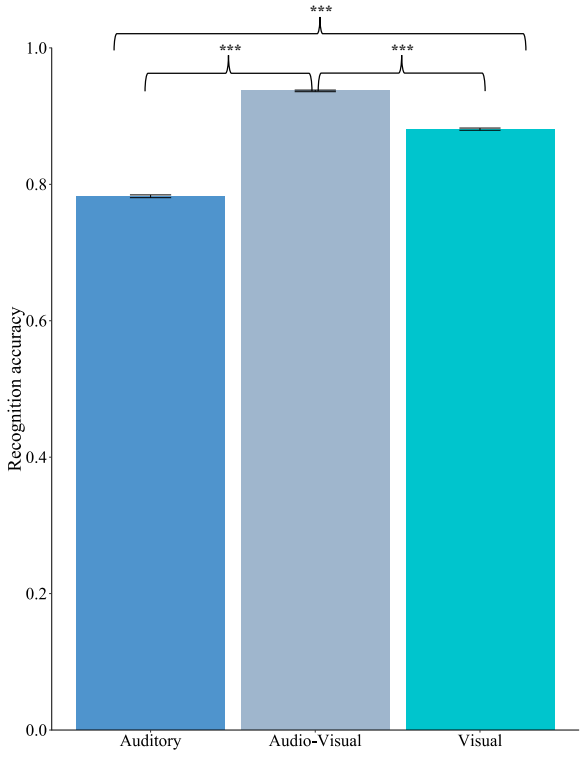


Figure 3 | Bar chart showing the recognition accuracy (RA) in the three conditions of stimulus presentation
Error bars represent the standard error. RA was significantly higher for the audio-visual presented stimuli than for the visual- or auditory stimuli. Accuracy scores were significantly higher for the visual- than for auditory condition.

Emotion specificity and modality [Aim 2]

The modality of stimulus presentation across fearful, disgusted and sad expressions significantly influenced participants' RA (*Friedman test*: $\chi^2_{(2)} = 400.47$, $p < 0.001$) and RT (*Friedman test*: $\chi^2_{(2)} = 208.77$, $p < 0.001$). Results comparing RA and RTs between

modalities for each emotion category showed that participants were significantly more accurate and faster at categorizing these emotions in the audio-visual than auditory modality (p's < 0.001; effect sizes for accuracy ranging from $0.813 < r < 0.852$ and for RT ranging from $0.422 < r < 0.760$). Although RA was significantly higher for disgust ($p < 0.001$; $r = 0.605$) and sad expressions ($p < 0.001$; $r = 0.417$) in the audio-visual than visual modality, the accuracy scores for fear did not significantly differ between these two modalities ($p = 1.00$; $r = 0.038$). Similarly, we observed no significant RT differences between the audio-visual and visual modality for these three emotions ($ps > 0.05$; $0.005 < r < 0.159$). While participants were significantly better at categorizing angry expressions in the voice than in the face ($p < 0.001$, $r = 0.492$), RTs did not differ significantly between these two modalities ($p = 1.00$, $r = 0.052$). In contrast, happy, disgusted, fearful, and sad expressions had significantly higher accuracy scores and faster RTs when they were presented visually than auditorily ($ps < 0.001$; $0.625 < r_{\text{Accuracy}} < 0.868$; $0.487 < r_{\text{RT}} < 0.816$). ***Table 3*** displays the test statistics for each modality and emotion category.

Table 3 | Recognition accuracy (RA) and reaction times (RTs) standardized *z-scores*, *p-values*, *95% confidence intervals* ($CI_{95\%}$) and *effect sizes* (*r*) for the comparisons between modalities by emotion categories

	Emotions	***RA***					***RT***				
		z	*p*	$CI_{95\%}$		*r*	*z*	*p*	$CI_{95\%}$		*r*
				LL	UL				LL	UL	
	Anger	13.71	<0.001	0.125	0.146	0.816	-8.645	<0.001	-0.299	-0.200	0.515
Audio-visual	Disgust	10.155	<0.001	0.104	0.125	0.605	0.550	1.00	-0.032	0.569	0.033
vs.	Fear	0.632	1.00	-0.000	0.021	0.038	2.677	0.134	0.019	0.126	0.159
Visual	Happiness	-2.820	0.087	-0.041	-0.000	0.168	3.397	0.012	0.018	0.072	0.202
	Sadness	6.995	<0.001	0.042	0.083	0.417	0.089	1.00	-0.044	0.051	0.005
	Neutral	9.547	<0.001	0.062	0.083	0.568	1.978	0.864	0.000	0.079	0.118
	Anger	10.579	<0.001	0.063	0.083	0.630	-6.736	<0.001	-0.302	-0.170	0.401
Audio-visual	Disgust	14.315	<0.001	0.250	0.271	0.852	-12.765	<0.001	-0.735	-0.562	0.760
vs.	Fear	13.646	<0.001	0.167	0.188	0.813	-9.653	<0.001	-0.526	-0.366	0.575
Auditory	Happiness	14.534	<0.001	0.188	0.208	0.865	-11.709	<0.001	-0.506	-0.373	0.697
	Sadness	13.858	<0.001	0.187	0.208	0.825	-7.087	<0.001	-0.359	-0.208	0.422
	Neutral	8.789	<0.001	0.062	0.083	0.523	-8.659	<0.001	-0384	-0.242	0.516
	Anger	8.268	<0.001	0.063	0.104	0.492	-0.865	1.00	-0.094	0.036	0.052
Auditory	Disgust	-10.50	<0.001	-0.187	-0.146	0.625	13.711	<0.001	0.597	0.746	0.816
vs.	Fear	-13.318	<0.001	-0.188	-0.167	0.793	12.113	<0.001	0.433	0.579	0.721
Visual	Happiness	-14.574	<0.001	-0.229	-0.188	0.868	13.51	<0.001	0.443	0.571	0.805
	Sadness	-11.603	<0.001	-0.187	-0.146	0.691	8.179	<0.001	0.232	0.370	0.487
	Neutral	0.941	1.00	-0.000	0.021	0.056	10.323	<0.001	0.295	0.420	0.615

Note: The differences in RA and RT between modalities by emotion categories were analyzed using *Wilcoxon-rank-sum test.* All *p-values* for RA and RT were for 18 comparisons (3 modalities * 6 emotions) Bonferroni corrected. *Positive z-scores* indicate that RA is higher and RTs longer for the first vs. second modality, whereas *negative z-scores* indicate that RA is lower and RTs shorter for the first vs. second modality.

Interplay of hormones, emotion recognition and reaction times [Aim 3]

Spearman's rank correlation coefficient between T1 and T2 for T was $rs = 0.79$ and $rs = 0.60$ for C. No significant associations between T or C and RA/RTs were found (p's $> .05$; correlation coefficients (r_s) close to zero; *Figure S1* in supplementary material illustrates the relationship between T or C and RA/RTs, also across all modalities). Similarly, there were no significant associations between T or C and RA/RTs for specific emotion categories (see *Table S2* in supplementary material). Logistic and linear models, however, showed that the interaction between testosterone and cortisol (TxC) significantly influenced participants' RA ($\chi^2_{(4)} = 46.30$, $p < 0.001$, $r = 0.022$) and RTs ($F_{(4,\ 121806)} = 8.26$, $p < 0.001$, $r = 0.016$). *Table S3* in supplementary material provides an overview on the model terms and the corresponding statistics for both RA and RTs. The odds ratio estimates for RA and the linear contrasts for the pattern of the differences in RTs for all combinations between T and C terciles showed that participants RA was significantly higher for T_{High}/C_{Low} and T_{Low}/C_{High}, but lower for T_{Middle}/C_{Low} or T_{Low}/C_{Middle}. RT's were

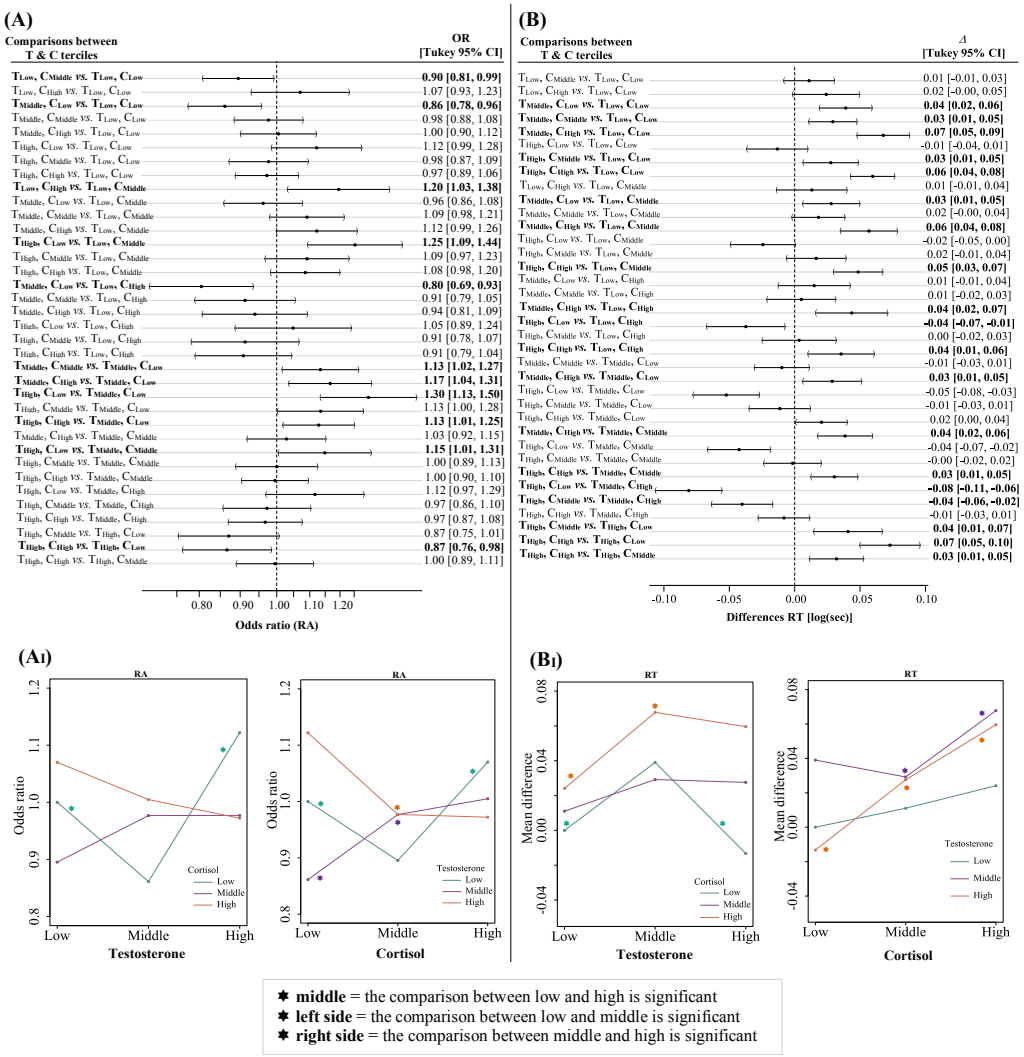


Figure 4 | Pairwise comparisons and conditional patterns of T and C terciles combinations for recognition accuracy (RA) and reaction time (RT)
The comparisons between hormone terciles for RA are illustrated in panel **(A)**, while the linear contrasts for the pattern of the differences in RT are illustrated in panel **(B)**. The significant combinations are highlighted in bold. The T pattern conditional under C and C pattern conditional under T for RA are shown in panel **(A_I)** and panel **(B_I)** for RT.
In panel **(A)** odds ratio for combination 1 (e.g., T_{High}/C_{High}) vs. combination 2 (e.g., T_{High}/C_{Low}) less than 1 indicate that the recognition probability for combination 2 (T_{High}/C_{Low}) is higher than for combination 1 (T_{High}/C_{High}), whereas values greater than 1 vice-versa. If the odds ratio of 1 is included in the confidence interval, the difference in the recognition probabilities is not significant. In panel **(B)** negative differences of RT for combination 1 (e.g., T_{High}/C_{High}) vs. combination 2 (e.g., T_{High}/C_{Low}) indicate that the RT for combination 2 (T_{High}/C_{High}) are longer than for combination 1 (T_{High}/C_{Low}), whereas positive differences vice-versa. If the difference of zero is included in the 95%CI, the difference in RT is not significant.
As it can be observed, for ***T conditional under C_{Low}*** and ***C conditional under T_{Low}*** there is a quadratic relationship [i.e., the accuracy decreases from low to middle T or C and then increases from middle to high T or C (see panel A_I); for ***T conditional under C_{Low}*** the RT increases from low to middle T and then decreases from middle to high T (see panel B_I)]. For ***C conditional under T_{High}*** the relationship is monotone [i.e., the accuracy decreases from low C to high C (see panel A_I); the RT increases from low C to high C (see panel B_I)].

shorter for T_{High}/C_{Low}, T_{Low}/C_{Low}, as well as, for T_{Low}/C_{Middle}. For the combinations T_{High}/C_{High} or T_{Middle}/C_{High} RTs were significantly longer. In ***Figure 4*** panels A, B display the corresponding statistics for all comparisons between T and C terciles, while panels A_I, B_I illustrate the conditional patterns.

4.4 Discussion

The main objective of the present study was to investigate whether males' RA is influenced by the modality of stimulus presentation in an explicit emotion recognition task. In

addition, we examined whether specific emotions are more quickly and accurately detected as a function of modality. Finally, we explored the effects of individual differences in T and C, as well as their interaction with RA and RTs.

Our results provide compelling evidence that RA is greatly improved when visual and audio information were jointly presented and that happy expressions were identified faster and with higher accuracy from faces than voices. Conversely, angry expressions were better recognized in voices than faces. Although no significant associations between single hormones (i.e., T or C) and RA or RTs were found, results showed that TxC interaction was significantly associated with both RA and RTs.

4.4.1 Emotion recognition performance as a function of modality and emotion category

Our data highlights that the audio-visual presentation of emotional expressions significantly contributes to the ease and efficiency with which others' emotions are recognized. This is in line with previous studies showing that the integration of auditorily and visually presented emotional information facilitates emotion recognition (e.g., Bänziger et al., 2009; Jessen et al., 2012; Paulmann & Pell, 2011), reflected in higher accuracy and faster RTs, especially for emotions such as disgust, fear (Collignon et al., 2008) and sadness (Kreifelts et al., 2007). One of the most noticeable differences between the present study and previous investigations was the presentation of several emotions and a neutral category (e.g., Collignon et al., 2008; De Gelder & Vroomen, 2000, included only two emotions) and the measurement of reaction time (e.g., not considered in Kreifelts et al., 2007, study). Yet, the facilitation effect concerning stimulus classification manifested for every single emotion category during the audio-visual modality in comparison to the auditory modality. In addition, RA in the audio-visual modality exceeded that of the visual modality for angry, disgusted, neutral and sad emotions, which indicates the comprehensive nature of this integration process. As shown by the present results there are some differences in the effectiveness, with which specific emotions are recognized from voices and faces. Similar to the results reported in a meta-analysis by Elfenbein and Ambady (2002a), anger was recognized better from voice than faces in our study, while better results for happiness were achieved from the visual compared to the auditory modality. This suggests that sensory modalities do not merely carry redundant information but rather, each may have certain specialized functions for the communication of emotions. Although the estimation of a visual threat (e.g., angry face) can be accurately predicted from close proximity, it has been shown that the louder, higher pitched sound of anger is particularly useful for both, proximal and distal spaces (see Ceravolo, Frühholz, & Grandjean, 2016, for details). As it is highly adaptive to recognize and react to a potential threat in the environment (Pichon, de Gelder, & Grezes, 2008), the accurate detection of anger might, therefore, rely more on the human auditory than visual system. Previous research on facial expression recognition has consistently reported that happy expressions are recognized more accurately and faster than other basic emotions (e.g., Nummenmaa & Calvo, 2015). Our data provide further support for these findings, but not for our prediction (1b) that emotions communicated by

the voice are recognized at higher rates of accuracy than in the visual channel. Nevertheless, it is possible that what determines the recognition advantage of happy faces is not so much their affect, but rather their perceptual and categorical distinctiveness from other emotional expressions (see, Calvo, Gutiérrez-García, Fernández-Martín, & Nummenmaa, 2014, for details) as well as their frequent occurrence in everyday social contexts, thus, tuning the visual system towards efficient recognition of these faces (Nummenmaa & Calvo, 2015). Moreover, it has been argued that physical feature extraction can occur instantaneously for facial expressions, while the interplay of acoustic cues over time occurs in a probabilistic manner (Juslin & Laukka, 2003) and thus, may not engage a similar process for vocal expressions (see Paulmann & Pell, 2011, for details). This could have strengthened the underlying knowledge about emotions leading to improved RA and RTs in the visual modality.

4.4.2 The interplay between hormones and ERA/RT

The available evidence regarding the relationship between T and males' emotion recognition ability is by no means clear-cut, making explicit claims about the direction of these effects impossible. The two predictions made in the present study were based on reported observations that T might have a negative influence on the recognition of emotions (Fujisawa & Shinohara, 2011; Rukavina et al., 2018) and that RTs of threat-related emotional expressions (i.e., angry, fear) would be much shorter with increasing levels of T (Derntl et al., 2009). To provide a more detailed picture of this association, we conducted an exploratory analysis for each modality and emotion category separately. In a similar fashion, we additionally analysed the effects of C. Similar to other reports in the literature, our data do not provide support for the influence of single steroid hormones (i.e., T or C) on RA or RTs (Derntl et al., 2009; Duesenberg et al., 2016). In contrast to the reported effect sizes or the significant effects between T and specific emotion categories (Derntl et al., 2009; Rukavina et al., 2018), the correlation coefficients for both hormones were small or close to zero across all modalities in our study. Despite our comparatively large sample, single hormones (i.e., T, C) did not appear to have an impact on RA and RTs in explicit emotion recognition tasks.

One assumption that has been put forth is that T and C do not act in isolation but rather interact to modulate complex social behaviours (Carré & Mehta, 2011). Following the dual-hormone hypothesis (Mehta & Josephs, 2010), we further explored whether the relationship between T and our response variables (i.e., RA and RT) is enhanced when C levels are low and attenuated when C levels are high. Similar to the obtained results in Dekkers (2018) meta-analysis, the overall effect size of T by cortisol interaction on RA and RT was significant but small in our study. Although our data support the dual-hormone hypothesis to some extent, they also showed that the interplay between T and C with RA or RTs is not as straightforward as one would expect. For instance, accuracy increased and RTs were shorter not only when T was high and C was low or vice-versa, but also when T and C were low. As our study is the first to account for the interaction between T and C on RA or RT, we cannot clearly provide explanations that might account for the observed

mixed-pattern of results. However, as previous research found that high T and stress (C) levels impair cognitive abilities (e.g., Gouchie & Kimura, 1991; Hänggi, 2004) and decrease performance (e.g., Dolcos, Wang, & Mather, 2014; Mehta, Wuehrmann, & Josephs, 2009), one would expect that with low levels of T and C, or with optimal levels of stress (i.e., eustress) but low T levels RA would increase in cognitive tasks. Since the pattern of the TxC interaction we found is unexpected and the effect size is small, we cannot rule out that it is a false-positive finding. Certainly, more work is needed to replicate our findings and to test these claims.

4.4.3 Strengths, limitations and future research

While our knowledge of how emotional information is integrated and recognized across channels is advancing steadily, the available literature, including the present study, is limited in a number of ways. In comparison to our study, most of the research mentioned above has evaluated a very small number of emotions (sometimes as few as two) and did not include a neutral baseline. Further, in some studies the audio material consisted of speech prosody (words, sentences). This opens up the possibility that the emotional tone of voice interacted with the affective value carried by the sentence's/word's semantic content. A related issue of past work is the use of emotional exemplars in conflict situations argued to be highly atypical of natural expressions of emotions (Paulmann & Pell, 2011). We addressed these issues by presenting emotion stimuli devoid of meaning (i.e., pseudo-words, pseudo-sentences and affect bursts) which always contained a congruent set of cues (i.e., encoder sex, stimulus time length) to express one of five basic emotions or a neutral state. We chose static faces to ensure our experimental conditions of stimulus presentation were compatible with the majority of prior literature. However, this format has been argued to be less ecologically valid (Krumhuber, Kappas, & Manstead, 2013; Recio et al., 2011). While this assumption is still subject to some controversy (see Dobs, Bülthoff, & Schultz, 2018, for details), future studies would benefit from using datasets of more naturalistic stimuli to further increase ecological validity.

As most of the previous research has focused on the associations between single hormones and facial emotion recognition, the present study uniquely contributes to the literature by providing a systematic examination of the influence of T, C and their interaction on RA and RT across different sensory modalities (i.e., auditory, visual and audio-visual). Although for C as well as for the interaction between T and C, the analyses were exploratory, they might prove of importance for researchers conducting work in this area to gain a more comprehensive understanding of when these effects emerge and when they do not. They may also yield a substantial theoretical payoff by enabling richer and more accurate predictions concerning the kind of outcomes tied to certain hormone level combinations.

The homogeneous characteristics of our sample (i.e., university students, narrow age range) may show patterns which do not hold for different sociodemographic subgroups. Given the increased focus on study replicability, future studies would benefit from combining datasets of different laboratories with similar outcome measures in order to reduce costs and increase the external validity, reliability and generalizability of findings. The

present study provided evidence for differences in both RA and RTs in the three conditions of stimulus presentation and potentially set the stage regarding the influence of TxC interaction on these two response variables. It would thus be worthwhile to expand on these findings and examine whether the same holds true for the other sex. This could be done, for instance, by investigating the interaction between oestradiol and cortisol with RA, as previous research showed that high oestradiol is associated with more externalizing behaviours (linked to emotion-recognition difficulties, see Chronaki et al., 2015), but only when cortisol was low (Tackett et al., 2015).

4.4.4 Conclusion

The findings of this study inform our current understanding with regard to the audio-visual integration of emotional signals among men by showing that audio-visual stimuli benefit RA over unimodal stimuli. They also explain inconsistencies in the past literature by highlighting that in explicit emotion recognition tasks voice-only expressions do not increase RA. Moreover, they replicate previous findings by establishing that for particular emotion categories RA and RTs vary as a function of modality. Crucially, our study contributes to a scientific domain that is currently reconsidering our understanding of the role hormones play for the recognition of emotions. It hereby paves the way for impactful future research, especially for the effects regarding TxC interaction with RA and RT.

Funding

This research was funded by Leibniz ScienceCampus "Primate Cognition"– Project number 6900199 *Leibniz-WissenschaftsCampus* and *Deutsche Forschungsgemeinschaft* (DFG, German Research Foundation) – Project number 254142454/GRK 2070.

Acknowledgments

The authors thank Edmund Henniges for technical support, Saskia Brueckner, Marc Koehler and Isabel Noethen for help with data acquisition and all individuals who participated in the research presented here.

4.5 Supplementary Material

Table S1 | Example of audio and visual stimuli matching for the audio-visual modality, duration, sex and emotion category

		Modality											
		Audio				*Visual*				*Audio-visual*			
Database	Stimulus	*Speaker ID*	*Speaker Sex*	*Emotions*	*Duration (s)*	*Actor ID*	*Actor Sex*	*Emotions*	*Duration (s)*	*Speaker/ Actor ID*	*Speaker/ Actor Sex*	*Emotions*	*Duration (s)*
Montreal Affective Voices	affect bursts	V01	M	Anger	1.142	F01	M	Anger	1.142	V01 F01	M	Anger	1.142
				Fear	1.0511			Fear	1.0511			Fear	1.0511
				Disgust	0.7613			Disgust	0.7613			Disgust	0.7613
				Happiness	1.7418			Happiness	1.7418			Happiness	1.7418
				Neutral	0.896			Neutral	0.896			Neutral	0.896
				Sadness	1.142			Sadness	1.142			Sadness	1.142
		V02	F	Anger	0.9492	F02	F	Anger	0.9492	V02 F02	F	Anger	0.9492
				Fear	0.6065			Fear	0.6065			Fear	0.6065
				Disgust	0.6284			Disgust	0.6284			Disgust	0.6284
				Happiness	1.5633			Happiness	1.5633			Happiness	1.5633
				Neutral	0.992			Neutral	0.992			Neutral	0.992
				Sadness	1.7795			Sadness	1.7795			Sadness	1.7795
Magdeburg Prosody Corpus	pseudo-word 1	V03	M	Anger	0.6792	F03	M	Anger	0.6792	V03 F03	M	Anger	0.6792
				Fear	1.0681			Fear	1.0681			Fear	1.0681
				Disgust	1.7066			Disgust	1.7066			Disgust	1.7066
				Happiness	0.5863			Happiness	0.5863			Happiness	0.5863
				Neutral	0.62			Neutral	0.62			Neutral	0.62
				Sadness	1.3003			Sadness	1.3003			Sadness	1.3003
	pseudo-word 1	V04	F	Anger	0.9375	F04	F	Anger	0.9375	V04 F04	F	Anger	0.9375
				Fear	1.0623			Fear	1.0623			Fear	1.0623
				Disgust	1.3119			Disgust	1.3119			Disgust	1.3119
				Happiness	0.9317			Happiness	0.9317			Happiness	0.9317
				Neutral	0.80			Neutral	0.80			Neutral	0.80
				Sadness	1.0877			Sadness	1.0877			Sadness	1.0877
	pseudo-word 2	V03	M	Anger	0.8853	F05	M	Anger	0.8853	V03 F05	M	Anger	0.8853
				Fear	0.7808			Fear	0.7808			Fear	0.7808
				Disgust	1.7879			Disgust	1.7879			Disgust	1.7879
				Happiness	0.6211			Happiness	0.6211			Happiness	0.6211
				Neutral	0.5562			Neutral	0.5562			Neutral	0.5562
				Sadness	1.1378			Sadness	1.1378			Sadness	1.1378
	pseudo-word 2	V04	F	Anger	1.2307	F06	F	Anger	1.2307	V04 F06	F	Anger	1.2307
				Fear	1.1494			Fear	1.1494			Fear	1.1494
				Disgust	1.5964			Disgust	1.5964			Disgust	1.5964
				Happiness	1.0101			Happiness	1.0101			Happiness	1.0101
				Neutral	0.9368			Neutral	0.9368			Neutral	0.9368
				Sadness	1.1461			Sadness	1.1461			Sadness	1.1461

Note: For affect bursts each actor voice was matched with an actor face for sex and stimulus length [e.g., anger uttered by the actor voice V01 with a duration of 1.142s was matched in the audio-visual modality with the model face F01 displaying the same emotion and was shown in the same time length as the auditory stimulus (i.e., 1.142s); F01 displaying anger was also played with a duration of 1.142s in the visual modality].
Each pseudo-word and pseudo-sentence spoken by a male and a female was matched by the face of a different model of the same sex.

Table S2 | Associations between T or C and recognition accuracy (RA)/reaction time (RT) by specific emotion categories for each- and across all 3 modalities of stimulus presentation.

		Testosterone				Cortisol			
Modalities	***Emotions***	***RA***		***RT***		***RA***		***RT***	
		p	r_s	*p*	r_s	*p*	r_s	*p*	r_s
Auditory	Anger	0.429	0.047	0.196	0.077	0.703	-0.023	0.264	0.067
	Disgust	0.346	0.056	0.767	-0.018	0.723	0.021	0.514	0.039
	Fear	0.252	-0.069	0.313	0.060	0.804	0.015	0.070	0.108
	Happiness	0.541	-0.037	0.260	0.067	0.842	-0.012	0.063	0.111
	Neutral	0.742	-0.020	0.951	0.004	0.195	-0.077	0.123	0.092
	Sadness	0.431	0.047	0.719	-0.022	0.793	-0.016	0.229	0.072
Visual	Anger	0.277	0.065	0.264	0.067	0.831	0.013	0.217	0.074
	Disgust	0.960	-0.003	0.951	0.004	0.066	-0.110	0.156	0.085
	Fear	0.723	-0.021	0.398	0.050	0.497	-0.041	0.332	0.058
	Happiness	0.224	-0.073	0.005	0.167	0.110	-0.096	0.073	0.107
	Neutral	0.818	0.014	0.211	0.075	0.745	0.019	0.454	0.045
	Sadness	0.234	-0.071	0.168	0.082	0.075	0.106	0.473	0.043
Audio-Visual	Anger	0.559	0.035	0.878	0.009	0.392	-0.051	0.316	0.060
	Disgust	0.297	0.062	0.824	-0.013	0.957	-0.003	0.182	0.080
	Fear	0.767	-0.018	0.703	0.023	0.879	-0.009	0.140	0.088
	Happiness	0.918	0.006	0.963	0.003	0.359	-0.055	0.128	0.091
	Neutral	0.489	0.041	0.630	0.029	0.631	0.029	0.156	0.085
	Sadness	0.870	-0.010	0.904	0.007	0.974	0.002	0.709	0.022
Across all 3 modalities	Anger	0.084	0.103	0.215	0.074	0.739	-0.020	0.089	0.102
	Disgust	0.250	0.069	0.735	-0.020	0.519	-0.039	0.248	0.069
	Fear	0.310	-0.061	0.370	0.054	0.815	-0.014	0.062	0.111
	Happiness	0.436	-0.047	0.119	0.093	0.515	-0.039	0.035	0.126
	Neutral	0.941	-0.004	0.794	0.016	0.659	0.026	0.204	0.076
	Sadness	0.890	0.008	0.744	0.020	0.595	0.032	0.273	0.066

Note: The tests were conducted using *Spearman's rank correlation*. As denoted by the correlation coefficients (r_s) the strength of the association between T or C and RA/RT for specific emotion categories was weak.

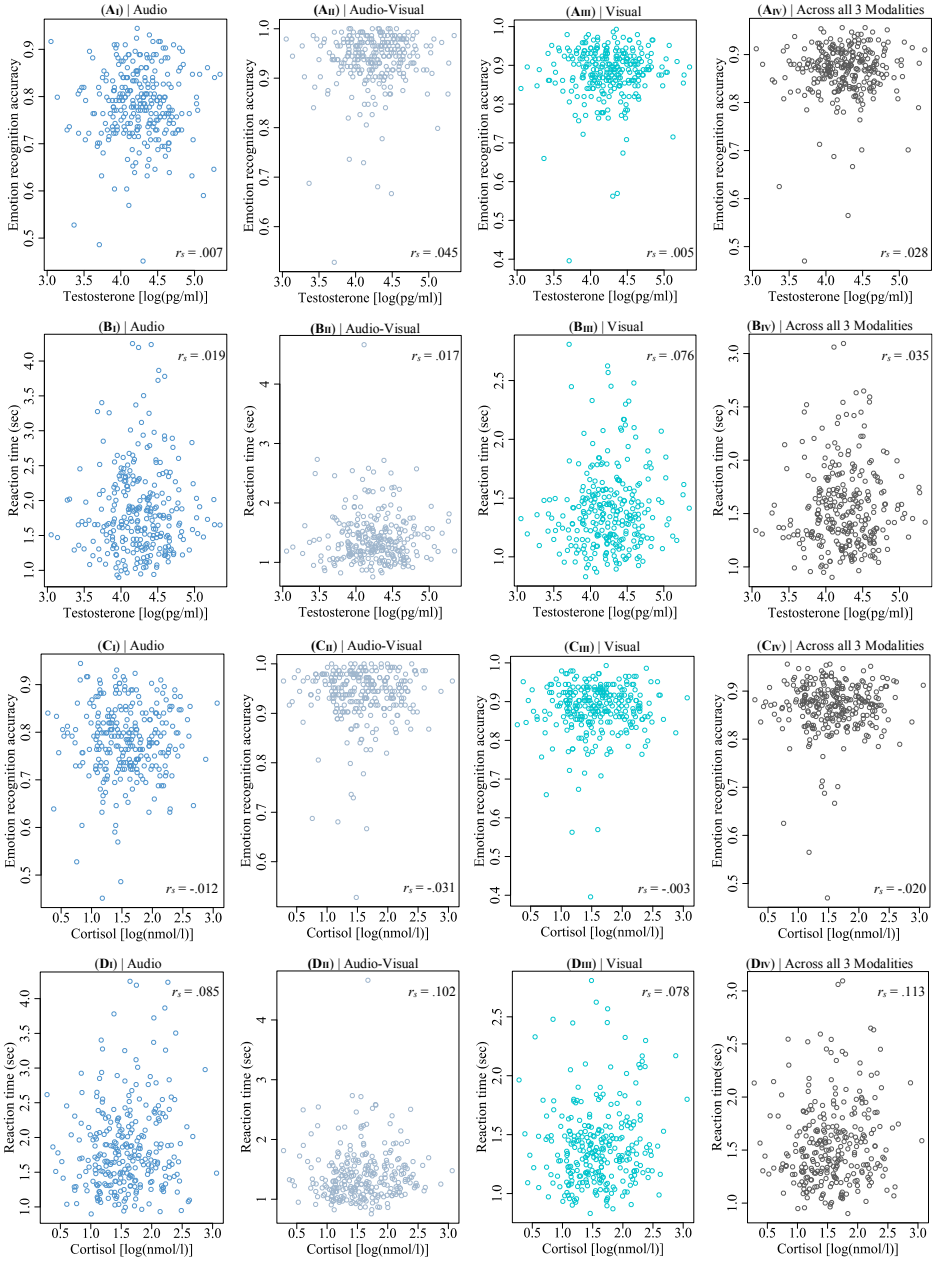


Figure S1 | Scatterplots for each- and across all modalities displaying the associations between T or C and emotion recognition accuracy (RA)/reaction time (RT)
The relationship between *Testosterone* & *RA* is shown in panels A_I to A_{IV}, *Testosterone* & *RT* in panels B_I to B_{IV}, *Cortisol* & *RA* in panels C_I to C_{IV} and *Cortisol* & *RT* in panels D_I to D_{IV}. *Spearman's rank correlation* tests showed no significant associations between these variables with *p*-values >.05 and weak rank correlation coefficients (r_s).

Table S3 | Logistic and linear models for T, C and TxC terciles

	Model terms	*Df*	*Deviance*	*Resid.Df*	*Resid.Dev*	*Pr(>Chi)*
	Null			121822	95628	
	Stimuli duration	1	137.60	121821	95491	< 0.001
Quasi-binomial logistic model	Emotions	5	2527.10	121816	92964	< 0.001
(*DV* = RA)	Modalities	2	4431.90	121814	88532	< 0.001
	Testosterone (T)	2	6.20	121812	88525	0.044
	Cortisol (C)	2	3.50	121810	88522	0.173
	TxC	4	46.30	121806	88476	< 0.001
	Model terms	***Df***	***SumSq***	***MeanSq***	***F-value***	***Pr(>F)***
	Stimuli duration	1	205	204.76	646.15	< 0.001
Linear model	Emotions	5	806	161.11	508.40	< 0.001
(*DV* = RT)	Modalities	2	1028	513.97	1621.87	< 0.001
	Testosterone (T)	2	30	15.21	48.01	< 0.001
	Cortisol (C)	2	35	17.73	55.94	< 0.001
	TxC	4	10	2.62	8.26	< 0.001
	Residuals	121806	38600	0.32		

Note: T and C were categorized using terciles. Effect sizes (*r*) for each variable in the quasi-binomial logistic model can be calculated as follows: **sqrt (Deviance/Null Resid.Dev)**. For the linear model the effect sizes (*r*) can be obtained by taking the **sqrt (SumSq/Total SumSq)**. To obtain η_p (see R package *lmSupport*), while for transformations to other effect sizes see https://www.psychometrica.de/effect_size.html.
DV = dependent variable; *RA* = recognition accuracy; *RT* = reaction time; *Resid.Df* = residual degrees of freedom; *Resid.Dev* = residual deviance; *SumSq* = sum of squares; *sqrt* = square root

Chapter 5

General Discussion

The overarching aim of this research project was to systematically examine factors that were argued to impact on individuals' ability to recognize basic emotions from voices and faces. Previous literature implicates gender, acoustic parameters, confidence judgments, hormonal fluctuations and modality of stimulus presentation as potential candidates that might influence performance accuracy in emotion recognition tasks. This chapter summarizes the main findings of the two conducted studies which were reported across three manuscripts. The three specific aims of this dissertation were to examine:

1. whether emotion recognition from vocal expressions differs as a function of listeners' and speakers' gender (*manuscript 1*)
2. whether the influence of vocal stimulus type and their related acoustic parameters influence emotion recognition and confidence ratings (*manuscript 2*)
3. whether males' ability to recognize emotions is influenced by the modality of stimulus presentation and hormonal fluctuations (*manuscript 3*)

This discussion will proceed to outline the specific points addressed for each of the three aims, followed by a brief summary of the main findings. These findings will be embedded in the existing literature and discussed alongside suggestions for future research.

Main findings

For the first aim (**1**), we specifically investigated:

- whether females perform better than males at recognizing emotions from the voice across all- and for each stimulus type, as well as, across all- and for each emotion category;
- if performance accuracy is higher for emotions spoken by female or male actors (i.e., across all- and for each stimulus type and each emotion category);
- potential interactions between listeners' and speakers' gender for the identification of vocal emotions (i.e., across all- and for each stimulus type).

We found no robust differences regarding the performance accuracy of recognizing emotions by listeners' gender. Also, no significant interactions between listeners' and speakers' gender were observed. An inspection of performance accuracy by emotion category and speakers' gender revealed a robust effect for disgust, indicating that listeners performed significantly better when this emotion was spoken by female- compared to male actors. Although significant, the pattern for the other emotion categories by speakers' gender was not as straightforward as one would expect. Contrary to the assumption that females are better in recognizing emotions and that some emotions are better understood when conveyed by a female speaker (e.g., Hall & Matsumoto, 2004), the reliability of emotion judgments, as shown by our data, is not systematically influenced by this factor and the related stereotypes of emotional expressivity. These findings are discussed below.

In contrast to the findings of a more recent study showing that females outperformed males in the recognition of disgust from facial expressions (Connolly et al., 2019), in our study, no significant differences for the recognition of *disgust* by listeners' gender were found. However, as indicated by our results, utterances spoken in a disgusted tone of voice had significantly higher accuracy scores for female compared to male actors in all types of stimuli except for pseudo-sentences. These findings pose the question of what could account for this female 'superiority' advantage in the recognition (cf., Connolly et al., 2019) and the expression of disgust. An interesting, if at present speculative perspective posited that the unique selection pressures faced by females, including immunosuppression during pregnancy and over the menstrual cycle, higher risk of contracting sexually-transmitted diseases and transferring them to their offspring, could count as potential factors that might enable females to recognize and express higher quality displays of disgust than males (e.g., Al-Shawaf, Conroy-Beam, Asao, & Buss, 2016; Fessler, Pillsworth, & Flamson, 2004; Fleischman, 2014). For instance, it has been suggested that in the realm of emotion communication, females are more likely to exaggerate their disgust expressions (i.e., because of educational purposes in the front of their offspring) than males, which are believed to downplay the expression of this emotion, potentially because they are more indifferent to harmful factors in their environment (Al-Shawaf et al., 2016; Al-Shawaf, Lewis, & Buss, 2018). Future research would benefit from a systematic investigation that accounts for both actors' and participants' gender, as well as, modalities of stimulus presentation in order to assess the robustness of this effect and, explores whether such fitness imperatives might shape the ability to express disgust in different ways across the sexes.

Our results also pose the question what might have led to the mixed patterns we observed in the recognition of other emotional expressions by speakers' gender. Examination of research on peoples gendered beliefs about emotion reveals that although people think that females are more emotional than males, they nonetheless believe that both sexes feel the same type and amount of emotions (e.g., Barrett, Robin, Pietromonaco, & Eyssell, 1998; LaFrance & Banaji, 1992). In other words, males and females are not thought to greatly differ in the extent to which they experience different emotions, but rather in the way they outwardly express their emotions to others (e.g., Fabes & Martin, 1991; Grossman & Wood, 1993). This stereotypical belief about the expression of emotion, as emphasized

by previous research, might lead to biased evaluations among participants. Therefore, when presented with an angry expression uttered/posed by a female, participants may choose other alternatives besides anger, which would lower the actress' apparent accuracy at expressing this emotion (e.g., Hall, 1984; Plant, Hyde, Keltner, & Devine, 2000). This explanation is quite puzzling, since such 'biases' in judgments could be accounted for by applying Wagner (1993) unbiased hit rate index of accuracy. In addition, previous research argued that these differences might be the result of the distinct social roles attributed to men and women. For example, females are thought to fulfill more care-taking roles and would hence show higher accuracy in the emotions involved in this role, such as happiness, fear and sadness (e.g., Grossman & Wood, 1993). Conversely, males are thought to feel and express anger more often, because this emotion is associated with their protective social role (e.g., Montagne, Kessels, Frigerio, de Haan, & Perrett, 2005). In contrast to these gender role theories and previously reported findings regarding the recognition and the expression of emotions, our results did not show a specific advantage for emotions for either males (e.g., anger) or females (e.g., happiness, fear, and sadness). Therefore, previous assumptions that locate the cause of these differences in either biological, social roles or sex-differentiated socialization pressures, could not be accounted for by our study.

So why do our findings diverge from what might be thought of as conventional wisdom (Timmers et al., 2003) or from other studies that do report a difference in females' and males' ability to recognize and express specific emotions? One possible explanation is that of publication bias in this field of research. This account was highlighted by A. E. Thompson and Voyer (2014) who conducted a meta-analysis on sex differences in emotion recognition and observed evidence for an excess of significant findings in the literature. Rather than the gender predispositions we addressed in our fist manuscript, another explanation of the above-mentioned accounts is that the strong heterogeneity of the stimulus types used in emotion recognition tasks could have led to the confusing pattern of results reported by previous investigations.

With this in mind, our second aim (**2**) was to examine:

- how much of the variance in recognition rates was explained by the acoustic attributes of emotive speech;
- if performance accuracy and confidence judgments were significantly higher/lower for certain types of vocal stimuli than others (e.g., affect bursts vs. neutral sentences) and, for specific emotions (e.g., angry vs. fear);
- whether confidence judgments were predicted by the correct recognition of vocal emotional expressions.

Research has long claimed that certain acoustic features, such as pitch, loudness, tempo or quality and their related parameters (e.g., fundamental frequency, jitter, shimmer, harmonics-to-noise ratio) drive the recognition of emotions from prosody and vocal bursts (e.g., Banse & Scherer, 1996; Sauter et al., 2010; K. R. Scherer & Bänziger, 2004). We extracted a baseline set of acoustic parameters from our stimuli datasets (see *manuscript 2*, for details) and employed two procedures to capture the psychophysical properties of

these measurements. First, discriminant analysis and random forest were implemented to determine whether this set of parameters provided sufficient information to successfully discriminate between stimuli from different emotional categories. Second, by employing a backward stepwise logistic regression analysis, we determined which of the acoustic predictors explained most of the deviance in listeners' recognition rates. Results showed high cross-validation estimates of accuracy for both classification methods, indicating that the stimuli contained detectable acoustic contrasts which helped listeners to differentiate the portrayed emotions and that most, if not all, parameters explained a significant amount of variance in listeners' recognition rates. This set of results corresponds to previous findings in the vocal emotion literature (e.g., Banse & Scherer, 1996; K. Hammerschmidt & Jürgens, 2007) and, in analogy, they parallel research on visual signals of emotions which reported that statistical classification methods can successfully discriminate facial expressions of different emotions on the basis of their pixel intensities (e.g., Calder, Burton, Miller, Young, & Akamatsu, 2001). Thus, for both vocal and facial modalities, it is possible to classify emotional expressions on the basis of basic perceptual features in a manner that models human performance.

Our results provide a clear indication that listeners were significantly more accurate and confident at judging emotions from vocal bursts than speech-embedded stimuli. This result corresponds to the idea that vocal bursts are more primitive and salient signals of emotion than speech-embedded vocalizations in an evolutionary sense (Krumhuber & Scherer, 2011). Since they do not require the dynamic spectral shaping caused by the rapid movements of the articulators (i.e., the tongue, jaw, lips and soft palate which shape the sound produced at the larynx), it has been suggested that vocal bursts resemble animal vocalizations more than they do spoken language (Krumhuber & Scherer, 2011; Scott, Sauter, & McGettigan, 2010). For instance, laughter has been described as more akin to modified breathing, involving inhalation as well as exhalation, than to speaking (K. J. Kohler, 2008). Therefore, it has been argued that the recognition of emotions from vocal bursts strongly depends on the preservation of acoustic information such as pitch or amplitude envelope variations (see Sauter, 2006, for details). Although voice quality, pitch, as well as, loudness, are important in the emotional inflection of spoken language (Banse & Scherer, 1996; Bänziger & Scherer, 2005), the recognition rates for speech-embedded emotions, as shown by our data and other studies (e.g., Hawk et al., 2009), are somewhat lower than they are for vocal bursts. So, what might account for these differences in recognition rates? While these differences may be due to the quality of our speech-embedded stimuli, the difference between emotional speech and vocal bursts may also reflect the fact that emotion in speech is overlaid on the speech signal. This would mean speech is somewhat more constrained in its emotional expression than are non-verbal vocalizations (Scott et al., 2010). Thus, there could be conflicts between the prosodic cues in sentence-level speech, which denote the emotional information, and those that cue linguistic information.

Similar to results reported in the literature (e.g., Belin et al., 2008; Hawk et al., 2009; Simon-Thomas, Sauter, Sinicropi-Yao, Abramson, & Keltner, 2007), our study revealed that vocal bursts proved to be highly effective means of expressing specific emotions, such

as disgust, happiness and sadness, in comparison to speech-embedded stimuli, with recognition accuracies above 80% (except for surprise). Results from cross-cultural studies corroborate these findings by reporting strong to moderate evidence for the universal recognizability of these emotions from vocal bursts (e.g., Cordaro et al., 2016; Cowen, Laukka, Elfenbein, Liu, & Keltner, 2019). Thus, one could ask whether there is a reason why these specific emotions are better recognized in this type of vocal stimuli? An argument that has been put forth is that vocal bursts are unique to some emotions (Goddard, 2014). For instance, laughter could be interpreted as a signal of happiness, crying as a signal of sadness, while interjections such as 'argh', 'eek' are typically indicative of disgust. Moreover, it has been suggested that people quite rarely vocalize disgust or surprise in the form of sentences (Banse & Scherer, 1996; Schaerlaeken & Grandjean, 2018). Since vocal bursts bear a heavy functional load in social interactions, as they are "so highly overlearned" and clearly attached to certain emotions (Goddard, 2014; K. R. Scherer, 1994), their accurate recognition might occur instantaneously and without conscious effort. In our study, anger had high accuracy scores (>80%) and confidence ratings for all speech-embedded stimuli. However, surprisingly, angry vocal bursts were the most difficult expressions to be recognized, after surprise. In a recent study, comparing the perceived authenticity of different vocal bursts corpora, Anikin and Lima (2018) showed that authentic vocalizations (e.g. anger, fear) differ from actor portrayals in a number of acoustic characteristics (showing a higher pitch, lower harmonicity, a variable spectral slope and amplitude). It is plausible that these acoustic characteristics of authenticity are hard-to-fake markers of a speaker's emotional state and thus signal a distinction between honest communication and a bluff. The Belin et al. (2008) stimuli, used in our study, received the lowest scores on perceived authenticity in the Anikin and Lima (2018) analysis. Thus, one could argue that the recognition pattern we observed for the angry vocal burst might be due to the characteristics of this database. In more general terms, our results regarding the efficiency of recognizing emotions from vocal bursts fit with two general interpretations. They may be explained by the innate psychological basicness of these signals (which might be universal to all humans), or by the psychological norms acquired by our listeners (which could have been shaped by culture).

Metacognition, the capacity to actively monitor and reflect upon one's own performance, has been argued to impact judgments of accuracy in emotion recognition tasks (Bègue, Adams, Stone, & Perez, 2019; Flavell, 1979; Kelly & Metcalfe, 2011; Koriat & Levy-Sadot, 1999). As indicated by our data, emotions that were confidently understood as such (e.g., happy) elicited correct and confident interpretations. This finding complements research on metacognition of facial expression recognition, showing that accurately assessing one's own performance results in improved emotion recognition (Kelly & Metcalfe, 2011). Together, these results suggest that the confidence with which an emotion expression is recognized may be just as important as decoding accuracy. However, more research is needed to uncover the underlying mechanisms of how individuals use this metacognitive knowledge. This could provide valuable insights on whether or not peoples' metacognition could be leveraged to improve emotion recognition (Kelly & Metcalfe, 2011) and might

serve as a point of departure for clinicians designing intervention strategies which help patients, lacking metacognitive awareness in their ability to recognize emotions, to develop the necessary skills to compensate for prosodic difficulties.

As shown by our first study, the effect sizes regarding sex differences in emotion recognition are comparably small, heterogeneity exists and the many factors that might lead to these differences are still unclear. Thus, in our second study we aimed to systematically analyze potential factors that might influence males' 'poor' performance accuracy in emotion recognition tasks.

Accordingly, for our third aim (**3**) we asked:

- if performance accuracy would be better for stimuli presented audio-visually, auditorily or visually (see *manuscript 3*, for hypotheses);
- whether the recognition of emotions is privileged to certain modalities (are some emotions more reliably recognized from the voice, face or when voice and face are jointly presented);
- if testosterone, cortisol, and their interaction influence performance accuracy and response times

Our results showed that performance accuracy is highest when visual and audio information are jointly presented. While happy expressions were identified faster and with higher accuracy from faces than voices, angry expressions were better recognized in voices than faces. Single hormone analysis of T and C did not reveal significant associations with performance accuracy and response times. The overall effect sizes of the testosterone by cortisol interaction with performance accuracy and response times was significant but small.

Given that much of our social interactions depend on the successful decoding of emotional information, it is critical to understand how we make use of different sources of emotional information and to identify whether we base emotional inferences on a particular hierarchy of information channels. Surprisingly, this topic has received relatively little empirical attention. Some test batteries were developed that allow insight into how multimodal stimulus processing may differ from unimodal processing in the recognition of emotions. For instance, Bänziger et al. (2009) developed the multimodal emotion recognition test (MERT) which contains dynamic emotional expressions from the auditory and visual modality alone or in combination. The authors reported higher emotion recognition when the face and voice are presented at the same time, as opposed to voice only. However, no such advantage was found between dynamic stimuli that contained face and voice information versus face information only, implying that facial expressions are more easily interpreted than vocal expressions. In addition, their data suggests that multimodal information does not necessarily lead to better performance accuracy in the recognition of emotional expressions. Our results, however, exemplify that as emotional channel availability increases, there is a corresponding increase in how accurately emotional displays are explicitly recognized (i.e., bimodal stimuli were recognized significantly better than unimodal stimuli). Thus, while there is evidence that emotions can be recognized fairly

well from only one channel in many instances (e.g., Laukka et al., 2016; Paulmann et al., 2008; K. R. Scherer et al., 2011) – also confirmed here, where we found that unbiased hit rates in the unimodal conditions ranged from .55 to .76 correct recognition for the auditory stimuli and .64 to .97 for the visual stimuli - our data establish that emotion recognition is facilitated by an enriched stimulus presentation. Our results are in line with other studies which showed that emotional judgments tend to improve when more than one source of congruent information about the intended emotion is available (e.g., Collignon et al., 2008; De Gelder & Vroomen, 2000; Kreifelts et al., 2007; Paulmann & Pell, 2011). Thus, it can be argued that more accurate recognition of bimodal versus unimodal stimuli provides indirect evidence for the integration of different information channels during emotional processing. It is reasonable to assume that emotional information channels need to be compared or integrated at some point to allow a holistic impression of the emotion being communicated. While our findings do not directly inform the nature of emotion integration, they are nonetheless consistent with the idea that emotion recognition processes incorporate all available information, possibly in an involuntary manner, leading to systematically higher accuracy rates. Interestingly, this process did not appear limited to the processing of emotional stimuli since we witnessed a similar advantage for neutral displays when presented audio-visually.

In addition to showing that the audio-visual presentation of stimuli facilitates emotion recognition, we examined whether particular channels are more effective for recognizing specific emotions. Overall, we noted that emotions presented in the visual modality were recognized more accurately than in the auditory channel. It is possible that these patterns reflect broad differences in how visual versus auditory information activate related emotion concepts during emotional communication. Specifically, one of the unique characteristics of emotional expressions conveyed through prosody is that they are inherently dynamic and their meaning unfolds in time (Juslin & Laukka, 2003), while the physical features which denote emotions in the facial channel (and similar to vocal bursts) can be processed instantaneously and without conscious effort. As assumed by previous investigations (e.g., Elfenbein & Ambady, 2002a) our results provide compelling evidence that angry expressions are recognized advantageously in the vocal channel, while happy expressions are better identified in faces than voices. This suggests that the salient features for recognizing discrete emotions are not always of equal value in the auditory and visual modality. For instance, Martinez and Du (2012) revealed a significant difference in recognition speed between different emotions. While happiness was the fastest emotion to be recognized, at 23-28ms after stimulus presentation, participants were about ten times slower to recognize anger. These results show that the recognition of emotions has evolved differently for distinct emotions, suggesting an adaptation to some evolutionary needs.

With respect to research on emotional processing and sex-/stress-related hormones, both psychological and neuroscientific studies often yield contradictory evidence, a phenomenon that has been termed a "replication crisis" (e.g., Maxwell, Lau, & Howard, 2015). Studies investigating the association between single hormones (testosterone; cortisol) and performance accuracy report either an increase, a decrease or null-effects. Researchers

have attributed these weak and inconsistent results to methodological limitations of studies. Another possible explanation that has been put forth by some investigators, that might account for the inconsistent findings, is that testosterone may act in concert with or in opposition to other hormones such as cortisol, to jointly regulate cognition and behavior (Sarkar et al., 2019). Growing evidence supports that testosterone-related behaviors, such as status seeking, risk-taking and aggression, are better explained by considering the interaction between cortisol and testosterone than by evaluating testosterone fluctuations in isolation (e.g., Mehta & Josephs, 2010; Mehta & Prasad, 2015). These findings are in line with the idea that environmental stress, as reflected by cortisol concentrations, would buffer or even halt the effect of testosterone on direct and indirect behaviors (see Viau, 2002, for details). Although subsequent studies do support this interpretation, they also extended it by showing that testosterone-cortisol interaction might take different forms depending on the specificity of the behavior and context. For instance, Welker et al. (2014) found that although both testosterone and cortisol were positively correlated with psychopathic traits in a non-clinical sample of young males, cortisol moderated the relationship between testosterone and psychopathy in males, such that high-testosterone and high-cortisol males reported the highest levels of psychopathy compared to the high-testosterone low-cortisol males. Therefore, one could argue that regardless of the form taken by this interaction, as shown by the patterns in our study, adding cortisol to the list of physiological modulators of testosterone release represents an important step towards a better understanding of how androgens shape social behavior and ultimately emotion recognition.

Limitations & Future research

As can be expected, the research conducted here has a number of limitations. The first limitation concerns an often-voiced criticism in the vocal emotion literature, namely the use of acted expressions. In other words, the internal and external validity of stimulus materials. Emotions portrayed in the laboratory have been argued to be exaggerated, stereotyped and 'un-natural' (e.g., Bachorowski & Owren, 1995; Barrett, 2006). However, as pointed out by some researchers addressing the question of internal validity of acted speech materials (e.g., Juslin & Laukka, 2001; Juslin & Scherer, 2005; Sauter et al., 2010), "natural" vocal expressions (i.e., emotions that are expressed outside or induced in the laboratory) do not necessarily represent more 'genuine' emotions, as they may also be contaminated by acting or social conventions. Addressing this issue, however, might improve methodology allowing researchers to combine different approaches and, therefore, represents an important point that deserves further empirical investigation. To address the issue of external validity, future studies could implement a metric that assesses the believability or authenticity of acted expressions. This could be done, for instance, by asking participants whether they thought the portrayal they heard/saw could be something they would hear/see in real-life. As some investigators argued that next to their limited ecological validity static expressions restrict our understanding of facial emotional expressions (Krumhuber et al., 2013), future research might utilize dynamic face stimuli. The absence of certain emotional categories within the databases, the fixed alternatives

of emotional categories listeners had to choose from, as well as, the complexity of some items (e.g., the perception of surprise might be interpreted as positive or negative) might have led to higher levels of cognitive load, which in turn might have affected, to some extent, performance accuracy. We decided to use a fixed-choice response format to compare our results with those of prior studies. This format, however, has been argued to be less ecologically valid (Russell, 1994) and may influence the level of response bias (Weijters, Cabooter, & Schillewaert, 2010). Despite the fact that we accounted for this potential bias by calculating the unbiased hit rates (Wagner, 1993), future studies could implement more valid alternatives, such as a slider (Kelly & Metcalfe, 2011) or visual analog scales (Young et al., 2017) as they may prove to be more sensitive when measuring individuals' emotion recognition ability. Also, the samples in our studies were limited to a university-educated population and included predominantly young adults, which may limit the generalizability of the findings to the wider population.

Conclusion

The results of the two manuscripts from Study I, expand previous research findings by showing that the magnitude between genders when decoding emotions from the voice is relatively small and, emotion recognition and confidence judgments are strongly influenced by vocal stimulus type and paralinguistic features of speech. Investigating the ways in which emotions can be expressed vocally, both in speech and in nonverbal expressions, contributes to a multimodal approach to emotional communication. While less empirical attention has been paid to how humans recognize emotions in the auditory modality from different types of stimuli, our data clearly underscore the rich significance of vocal bursts. The results from Study 2 establish that emotion recognition is more successful when several information channels are simultaneously present, leading to the assumption that emotional information in each channel is somehow integrated to form a unified impression about a interlocutors' emotion. The fact that multiple, congruent channels enhance recognition processes may be explained by increased activation of emotion-related knowledge or "emotion concepts" which are used during emotional communication, and in the formation of social impressions. A main implication of this study regards the interaction between testosterone and cortisol by showing the added value of their interaction, as compared to only studying main effects. Crucially, the results from the two studies actively contributes to a scientific domain that is currently re-writing our understanding of the role various factors play for the recognition of emotions. It hereby paves the way for impactful future research, especially for the effects regarding TxC interaction on emotion recognition accuracy.

References

Abele, A. (1985). Thinking about thinking: Causal, evaluative and finalistic cognitions about social situations. *European Journal of Social Psychology*, *15*(3), 315–332.

Abelson, R. P. (1985). A variance explanation paradox: When a little is a lot. *Psychological Bulletin*, *97*(1), 129–133.

Adams, S. M. (2012). Decoding nonverbal expressions of emotion of men and women. *Modern Psychological Studies*, *18*(1), 6–17.

Al-Shawaf, L., Conroy-Beam, D., Asao, K., & Buss, D. M. (2016). Human emotions: An evolutionary psychological perspective. *Emotion Review*, *8*(2), 173–186.

Al-Shawaf, L., Lewis, D. M., & Buss, D. M. (2018). Sex differences in disgust: why are women more easily disgusted than men? *Emotion review*, *10*(2), 149–160.

Altrov, R., Pajupuu, H., & Pajupuu, J. (2013). The role of empathy in the recognition of vocal emotions. In *Interspeech* (pp. 1341–1344).

Ambady, N., & Rosenthal, R. (1998). Nonverbal communication. *Encyclopedia of Mental Health*, *2*, 775–782.

Anikin, A., & Lima, C. F. (2018). Perceptual and acoustic differences between authentic and acted nonverbal emotional vocalizations. *The Quarterly Journal of Experimental Psychology*, *71*, 622–641.

Apple, W., & Hecht, K. (1982). Speaking emotionally: The relation between verbal and vocal communication of affect. *Journal of Personality and Social Psychology*, *42*(5), 864–875.

Arioli, M., Crespi, C., & Canessa, N. (2018). Social cognition through the lens of cognitive and clinical neuroscience. *BioMed Research International*, 1–18.

Arnal, L. H., Flinker, A., Kleinschmidt, A., Giraud, A.-L., & Poeppel, D. (2015). Human screams occupy a privileged niche in the communication soundscape. *Current Biology*, *25*(15), 2051–2056.

Association, W. M. (2013). World medical association declaration of helsinki: ethical principles for medical research involving human subjects. *Journal of the American Medical Association (JAMA)*, *310*(20), 2191–2194.

Azul, D. (2013). How do voices become gendered? a critical examination of everyday and medical constructions of the relationship between voice, sex, and gender identity. In *Challenging popular myths of sex, gender and biology* (pp. 77–88). Springer.

Babchuk, W. A., Hames, R. B., & Thompson, R. A. (1985). Sex differences in the recognition of infant facial expressions of emotion: The primary caretaker hypothesis. *Ethology and Sociobiology*, *6*(2), 89–101.

Bachorowski, J.-A., & Owren, M. J. (1995). Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context. *Psychological science*, *6*(4), 219–224.

Bąk, H. (2016). The state of emotional prosody research—a meta-analysis. In H. Bąk (Ed.), *Emotional prosody processing for non-native english speakers* (pp. 79–115). Springer.

Balconi, M. (2010). The neuropsychology of nonverbal communication: the facial expressions of emotions. In M. Balconi (Ed.), *Neuropsychology of communication* (pp. 177–202). Springer.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614–636.

Bänziger, T., Grandjean, D., & Scherer, K. R. (2009). Emotion recognition from expressions in face, voice, and body: the multimodal emotion recognition test (MERT). *Emotion*, *9*(5), 691–704.

Bänziger, T., & Scherer, K. R. (2005). The role of intonation in emotional expressions. *Speech communication*, *46*(3-4), 252–267.

Barrett, L. F. (2006). Are emotions natural kinds? *Perspectives on Psychological Science*, *1*(1), 28–58.

Barrett, L. F., & Bliss-Moreau, E. (2009). She's emotional. he's having a bad day: Attributional explanations for emotion stereotypes. *Emotion*, *9*(5), 649–658.

Barrett, L. F., Lindquist, K. A., & Gendron, M. (2007). Language as context for the perception of emotion. *Trends in Cognitive Sciences*, *11*(8), 327–332.

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science*, *20*(5), 286–290.

Barrett, L. F., Robin, L., Pietromonaco, P. R., & Eyssell, K. M. (1998). Are women the "more emotional" sex? evidence from emotional experiences in social context. *Cognition & Emotion*, *12*(4), 555–578.

Batson, C. D. (2009). These things called empathy: eight related but distinct phenomena. In *The social neuroscience of empathy* (pp. 3–16). Cambridge, MA: MIT Press.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370.

Beaupré, M. G., & Hess, U. (2006). An ingroup advantage for confidence in emotion recognition judgments: The moderating effect of familiarity with the expressions of outgroup members. *Personality and Social Psychology Bulletin*, *32*(1), 16–26.

Bègue, I., Adams, C., Stone, J., & Perez, D. L. (2019). Structural alterations in functional neurological disorder and related conditions: A software and hardware problem? *NeuroImage: Clinical*, 101798.

Belin, P. (2006). Voice processing in human and non-human primates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*(1476), 2091–2107.

Belin, P., Bestelmeyer, P. E., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology*, *102*(4), 711–725.

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The montreal affective voices: a

validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods*, *40*(2), 531–539.

Ben-David, B. M., Multani, N., Shakuf, V., Rudzicz, F., & van Lieshout, P. H. (2016). Prosody and semantics are separate but not separable channels in the perception of emotional speech: Test for rating of emotions in speech. *Journal of Speech, Language and Hearing Research*, *59*(1), 72–89.

Bestelmeyer, P. E., Kotz, S. A., & Belin, P. (2017). Effects of emotional valence and arousal on the voice perception network. *Social Cognitive and Affective Neuroscience*, *12*(8), 1351–1358.

Boersma, P. (2001). *Praat. a system for doing phonetics by computer.*

Bonebright, T. L., Thompson, J. L., & Leger, D. W. (1996). Gender stereotypes in the expression and perception of vocal affect. *Sex Roles*, *34*(5-6), 429–445.

Bostanov, V., & Kotchoubey, B. (2004). Recognition of affective prosody: Continuous wavelet measures of event-related brain potentials to emotional exclamations. *Psychophysiology*, *41*(2), 259–268.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Breyer, B., & Bluemke, M. (2016). Deutsche version der positive and negative affect schedule PANAS (GESIS panel). In *Zusammenstellung Sozialwissenschaftlicher Items und Skalen (Mannheim: GESIS)* (Vol. 10).

Briton, N. J., & Hall, J. A. (1995). Beliefs about female and male nonverbal communication. *Sex Roles*, *32*(1-2), 79–90.

Brody, L. R. (1997). Gender and emotion: Beyond stereotypes. *Journal of Social Issues*, *53*(2), 369–393.

Bryant, G. A., Fessler, D. M., Fusaroli, R., Clint, E., Aarøe, L., Apicella, C. L., ... others (2016). Detecting affiliation in colaughter across 24 societies. *Proceedings of the National Academy of Sciences*, *113*(17), 4682–4687.

Bublatzky, F., Gerdes, A., White, A. J., Riemer, M., & Alpers, G. W. (2014). Social and emotional relevance in face processing: happy faces of future interaction partners enhance the late positive potential. *Frontiers in Human Neuroscience*, *8*, 1–10.

Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. F., & Weiss, B. (2005). A database of german emotional speech. In *European conference on speech and language processing (eurospeech) (lisbon)* (p. 1517-1520).

Burton, L., Bensimon, E., Allimant, J. M., Kinsman, R., Levin, A., Kovacs, L., ... Bahrami, J. (2013). Relationship of prosody perception to personality and aggression. *Current Psychology*, *32*(3), 275–280.

Cacioppo, J. (2013). Why are we emotional? In *Discovering psychology: the science of mind, briefer version* (pp. 250–272). Belmont, USA: Wadsworth Cengage Learning.

Calder, A. J., Burton, A. M., Miller, P., Young, A. W., & Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vision research*, *41*(9), 1179–1208.

Calder, A. J., Young, A. W., Perrett, D. I., Etcoff, N. L., & Rowland, D. (1996). Categorical perception of morphed facial expressions. *Visual Cognition*, *3*(2), 81–118.

Calvo, M. G., Gutiérrez-García, A., Fernández-Martín, A., & Nummenmaa, L. (2014).

Recognition of facial expressions of emotion is related to their frequency in everyday life. *Journal of Nonverbal Behavior*, *38*(4), 549–567.

Carré, J. M., & Mehta, P. H. (2011). Importance of considering testosterone–cortisol interactions in predicting human aggression and dominance. *Aggressive Behavior*, *37*(6), 489–491.

Cassels, T. G., & Birch, S. A. (2014). Comparisons of an open-ended vs. forced-choice 'mind reading'task: Implications for measuring perspective-taking and emotion recognition. *PloS One*, *9*(12), e93653.

Castro, S. L., & Lima, C. F. (2010). Recognizing emotions in spoken language: A validated set of portuguese sentences and pseudosentences for research on emotional prosody. *Behavior Research Methods*, *42*(1), 74–81.

Ceravolo, L., Frühholz, S., & Grandjean, D. (2016). Proximal vocal threat recruits the right voice-sensitive auditory cortex. *Social Cognitive and Affective Neuroscience*, *11*(5), 793–802.

Chaplin, T. M. (2015). Gender and emotion expression: A developmental contextual perspective. *Emotion Review*, *7*(1), 14–21.

Chronaki, G., Garner, M., Hadwin, J. A., Thompson, M. J., Chin, C. Y., & Sonuga-Barke, E. J. (2015). Emotion-recognition abilities and behavior problem dimensions in preschoolers: evidence for a specific role for childhood hyperactivity. *Child Neuropsychology*, *21*(1), 25–40.

Chronaki, G., Wigelsworth, M., Pell, M. D., & Kotz, S. A. (2018). The development of cross-cultural recognition of vocal emotion during childhood and adolescence. *Scientific Reports*, *8*(1), 8659.

Clore, G. L., Schwarz, N., & Conway, M. (1994). Affective causes and consequences of social information processing. In *Handbook of social cognition: Basic processes; applications* (pp. 323–417). Hillsdale: Lawrence Erlbaum Associates, Inc.

Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., & Lepore, F. (2008). Audio-visual integration of emotion expression. *Brain Research*, *1242*, 126–135.

Collignon, O., Girard, S., Gosselin, F., Saint-Amour, D., Lepore, F., & Lassonde, M. (2010). Women process multisensory emotion expressions more efficiently than men. *Neuropsychologia*, *48*(1), 220–225.

Connolly, H. L., Lefevre, C. E., Young, A. W., & Lewis, G. J. (2019). Sex differences in emotion recognition: Evidence for a small overall female superiority on facial disgust. *Emotion*, *19*(3), 455–464.

Cordaro, D. T., Keltner, D., Tshering, S., Wangchuk, D., & Flynn, L. M. (2016). The voice conveys emotion in ten globalized cultures and one remote village in bhutan. *Emotion*, *16*(1), 117–128.

Cordaro, D. T., Sun, R., Keltner, D., Kamble, S., Huddar, N., & McNeil, G. (2018). Universals and cultural variations in 22 emotional expressions across five cultures. *Emotion*, *18*(1), 75–93.

Cornew, L., Carver, L., & Love, T. (2010). There's more to emotion than meets the eye:

A processing bias for neutral content in the domain of emotional prosody. *Cognition and Emotion*, *24*(7), 1133–1152.

Côté, S., & Hideg, I. (2011). The ability to influence others via emotion displays: A new dimension of emotional intelligence. *Organizational Psychology Review*, *1*(1), 53–71.

Cowen, A. S., Elfenbein, H. A., Laukka, P., & Keltner, D. (2018). Mapping 24 emotions conveyed by brief human vocalization. *American Psychologist*.

Cowen, A. S., Laukka, P., Elfenbein, H. A., Liu, R., & Keltner, D. (2019). The primacy of categories in the recognition of 12 emotions in speech prosody across two cultures. *Nature Human Behaviour*, 1.

Cuff, B. M., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: a review of the concept. *Emotion Review*, *8*(2), 144–153.

Cunningham, M. R. (1977). Personality and the structure of the nonverbal communication of emotion. *Journal of Personality*, *45*(4), 564–584.

Damasio, A. R., Grabowski, T. J., Bechara, A., Damasio, H., Ponto, L. L., Parvizi, J., & Hichwa, R. D. (2000). Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nature Neuroscience*, *3*(10), 1049–1056.

Danner, D., Rammstedt, B., Bluemke, M., Treiber, L., Berres, S., Soto, C., & John, O. (2016). Die deutsche version des big five inventory 2 (BFI-2). In *Zusammenstellung sozialwissenschaftlicher items und skalen*.

Darwin, C. (1872). The expression of emotion in animals and man. *London, England: Murray*.

Davis, E., Greenberger, E., Charles, S., Chen, C., Zhao, L., & Dong, Q. (2012). Emotion experience and regulation in china and the united states: how do culture and gender shape emotion responding? *International Journal of Psychology*, *47*(3), 230–239.

De Gelder, B. (2016). Gender, culture and context differences in recognition of bodily expressions. In B. De Gelder (Ed.), *Emotions and the body* (pp. 163–191). Oxford University Press.

de Gelder, B., Teunisse, J.-P., & Benson, P. J. (1997). Categorical perception of facial expressions: Categories and their internal structure. *Cognition & Emotion*, *11*(1), 1–23.

De Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition & Emotion*, *14*(3), 289–311.

Dekkers, M. T. (2018). A meta-analytical evaluation of the dual-hormone hypothesis: Does cortisol moderate the relationship between testosterone and status, dominance, risk taking, aggression, and psychopathy? *Neuroscience & Biobehavioral Reviews*(96), 250–271.

Demenescu, L. R., Kato, Y., & Mathiak, K. (2015). Neural processing of emotional prosody across the adult lifespan. *BioMed Research International*, *1*(9).

Derntl, B., Kryspin-Exner, I., Fernbach, E., Moser, E., & Habel, U. (2008). Emotion recognition accuracy in healthy young females is associated with cycle phase. *Hormones and Behavior*, *53*(1), 90–95.

Derntl, B., Windischberger, C., Robinson, S., Kryspin-Exner, I., Gur, R. C., Moser, E., &

Habel, U. (2009). Amygdala activity to fear and anger in healthy young males is associated with testosterone. *Psychoneuroendocrinology*, *34*(5), 687–693.

Derntl, B., Windischberger, C., Robinson, S., Lamplmayr, E., Kryspin-Exner, I., Gur, R. C., ... Habel, U. (2008). Facial emotion recognition and amygdala activation are associated with menstrual cycle phase. *Psychoneuroendocrinology*, *33*(8), 1031–1040.

De Waal, F. B. (2003). Darwin's legacy and the study of primate visual communication. *Annals of the New York Academy of Sciences*, *1000*(1), 7–31.

Dobs, K., Bülthoff, I., & Schultz, J. (2018). Use and usefulness of dynamic face stimuli for face perception studies–a review of behavioral findings and methodology. *Frontiers in Psychology*, *9*, 1355.

Dolcos, F., Wang, L., & Mather, M. (2014). Current research and emerging directions in emotion-cognition interactions. *Frontiers in Integrative Neuroscience*, *8*, 83.

Duesenberg, M., Weber, J., Schulze, L., Schaeuffele, C., Roepke, S., Hellmann-Regen, J., ... Wingenfeld, K. (2016). Does cortisol modulate emotion recognition and empathy? *Psychoneuroendocrinology*, *66*, 221–227.

Dunlosky, J., & Metcalfe, J. (2008). Confidence judgements. In *Metacognition* (pp. 118–139). Washington D.C.: Sage Publications.

Dziobek, I., Rogers, K., Fleck, S., Bahnemann, M., Heekeren, H. R., Wolf, O. T., & Convit, A. (2008). Dissociation of cognitive and emotional empathy in adults with asperger syndrome using the multifaceted empathy test (MET). *Journal of Autism and Developmental Disorders*, *38*(3), 464–473.

Eagly, A. H., & Wood, W. (1999). The origins of sex differences in human behavior: Evolved dispositions versus social roles. *American psychologist*, *54*(6), 408–423.

Eimer, M., & Holmes, A. (2007). Event-related brain potential correlates of emotional face processing. *Neuropsychologia*, *45*(1), 15–31.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*(3-4), 169–200.

Elfenbein, H. A., & Ambady, N. (2002a). Is there an in-group advantage in emotion recognition? *Psychological Bulletin*, *128*, 243–249.

Elfenbein, H. A., & Ambady, N. (2002b). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological Bulletin*, *128*, 203–235.

Ellsworth, P., & Scherer, K. R. (2001). Appraisal processes in emotion. In *Handbook of affective sciences* (pp. 572–595). New York: Oxford University Press.

Eyben, F., Batliner, A., & Schuller, B. (2010). Towards a standard set of acoustic features for the processing of emotion in speech. In *Proceedings of meetings on acoustics 159asa* (Vol. 9, pp. 2–12).

Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., André, E., Busso, C., ... others (2016). The geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing. *IEEE Transactions on Affective Computing*, *7*(2), 190–202.

Fabes, R. A., & Martin, C. L. (1991). Gender and age stereotypes of emotionality. *Personality and social psychology bulletin*, *17*(5), 532–540.

Feeney, J., Gaffney, P., & O'Mara, S. M. (2012). Age and cortisol levels modulate judgment of positive and negative facial expressions. *Psychoneuroendocrinology*, *37*(6), 827–835.

Fehr, B., & Russell, J. A. (1984). Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology: General*, *113*(3), 464–486.

Fessler, D. M., Pillsworth, E. G., & Flamson, T. J. (2004). Angry men and disgusted women: An evolutionary approach to the influence of emotions on risk taking. *Organizational behavior and human decision processes*, *95*(1), 107–123.

Fischer, A., & Evers, C. (2013). The social basis of emotion in men and women. In *The sage handbook of gender and psychology* (pp. 183–198). Sage Publications.

Fischer, A. H., Kret, M. E., & Broekens, J. (2018). Gender differences in emotion perception and self-reported emotional intelligence: A test of the emotion sensitivity hypothesis. *PloS One*, *13*(1), e0190712.

Fischer, A. H., & LaFrance, M. (2015). What drives the smile and the tear: Why women are more emotionally expressive than men. *Emotion Review*, *7*(1), 22–29.

Fischer, A. H., & Manstead, A. S. (2008). Social functions of emotion. In M. Lewis, J. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (Vol. 3, pp. 456–468).

Fischer, J., Metz, M., Cheney, D. L., & Seyfarth, R. M. (2001). Baboon responses to graded bark variants. *Animal Behaviour*, *61*(5), 925–931.

Fischer, J., & Price, T. (2017). Meaning, intention, and inference in primate vocal communication. *Neuroscience & Biobehavioral Reviews*, *82*, 22-31.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist*, *34*(10), 906.

Fleischman, D. S. (2014). Women's disgust adaptations. In *Evolutionary perspectives on human sexual psychology and behavior* (pp. 277–296). Springer.

Fleming, A. S., Corter, C., Stallings, J., & Steiner, M. (2002). Testosterone and prolactin are associated with emotional responses to infant cries in new fathers. *Hormones and Behavior*, *42*(4), 399–413.

Forni-Santos, L., & Osório, F. L. (2015). Influence of gender in the recognition of basic facial expressions: A critical literature review. *World Journal of Psychiatry*, *5*(3), 342–351.

Fraccaro, P. J., O'Connor, J. J., Re, D. E., Jones, B. C., DeBruine, L. M., & Feinberg, D. R. (2013). Faking it: deliberately altered voice pitch and vocal attractiveness. *Animal Behaviour*, *85*(1), 127–136.

Frank, M. G., & Stennett, J. (2001). The forced-choice paradigm and the perception of facial expressions of emotion. *Journal of Personality and Social Psychology*, *80*(1), 75–85.

Frijda, N., Scherer, K., & Sander, D. (2009). Emotion definitions (psychological perspectives). In *The oxford companion to emotion and the affective sciences* (1st ed., pp. 142–144). Oxford University Press.

Frith, C. D., & Frith, U. (2007). Social cognition in humans. *Current Biology*, *17*(16),

R724–R732.

Fujisawa, T. X., & Shinohara, K. (2011). Sex differences in the recognition of emotional prosody in late childhood and adolescence. *The Journal of Physiological Sciences*, *61*(5), 429–435.

Gaddy, M. A., & Ingram, R. E. (2014). A meta-analytic review of mood-congruent implicit memory in depressed mood. *Clinical Psychology Review*, *34*(5), 402–416.

George, N. (2013). The facial expression of emotions. In *The cambridge handbook of human affective neuroscience* (pp. 171–197). New York: Cambridge University Press.

Gery, I., Miljkovitch, R., Berthoz, S., & Soussignan, R. (2009). Empathy and recognition of facial expressions of emotion in sex offenders, non-sex offenders and normal controls. *Psychiatry Research*, *165*(3), 252–262.

Giardino, J., Gonzalez, A., Steiner, M., & Fleming, A. S. (2008). Effects of motherhood on physiological and subjective responses to infant cries in teenage mothers: a comparison with non-mothers and adult mothers. *Hormones and Behavior*, *53*(1), 149–158.

Gignell, M., Hornung, J., & Derntl, B. (2019). Emotional processing and sex hormones. In O. C. Schultheiss & P. H. Mehta (Eds.), *Routledge international handbook of social neuroendocrinology* (1st ed., pp. 403–419). Abingdon, UK: Routledge.

Goddard, C. (2014). Interjections and emotion (with special reference to "surprise" and "disgust"). *Emotion Review*, *6*(1), 53–63.

Goldman, A. I., & Sripada, C. S. (2005). Simulationist models of face-based emotion recognition. *Cognition*, *94*(3), 193–213.

Gouchie, C., & Kimura, D. (1991). The relationship between testosterone levels and cognitive ability patterns. *Psychoneuroendocrinology*, *16*(4), 323–334.

Goudbeek, M., & Scherer, K. (2010). Beyond arousal: Valence and potency/control cues in the vocal expression of emotion. *The Journal of the Acoustical Society of America*, *128*(3), 1322–1336.

Gray, J. (1992). *Men are from mars, women are from venus: A practical guide for improving communication and getting what you want in a relationship.* HarperCollins, New York.

Gregory Jr, S. W., & Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *Journal of Personality and Social Psychology*, *70*(6), 1231–1240.

Grossman, M., & Wood, W. (1993). Sex differences in intensity of emotional experience: a social role interpretation. *Journal of personality and social psychology*, *65*(5), 1010.

Hakamata, Y., Komi, S., Moriguchi, Y., Izawa, S., Motomura, Y., Sato, E., ... others (2017). Amygdala-centred functional connectivity affects daily cortisol concentrations: a putative link with anxiety. *Scientific Reports*, *7*(1), 8313.

Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin*, *85*(4), 845–857.

Hall, J. A. (1984). *Nonverbal sex differences: communication accuracy and expressive style.* Johns Hopkins University Press.

Hall, J. A. (2006). Nonverbal behavior, status, and gender: How do we understand their relations? *Psychology of Women Quarterly*, *30*(4), 384–391.

Hall, J. A., Carter, J. D., & Horgan, T. G. (2000). Gender differences in nonverbal communication of emotion. In A. H. Fischer (Ed.), *Gender and emotion: Social psychological perspectives* (pp. 97–117). Cambridge: Cambridge University Press.

Hall, J. A., & Matsumoto, D. (2004). Gender differences in judgments of multiple emotions from facial expressions. *Emotion*, *4*(2), 201–206.

Hamilton, D. L., & Huffman, L. J. (1971). Generality of impression-formation processes for evaluative and nonevaluative judgments. *Journal of Personality and Social Psychology*, *20*(2), 200–207.

Hamilton, D. L., & Zanna, M. P. (1972). Differential weighting of favorable and unfavorable attributes in impressions of personality. *Journal of Experimental Research in Personality*(6), 204–212.

Hammerschmidt, K., & Fischer, J. (2019). Baboon vocal repertoires and the evolution of primate vocal diversity. *Journal of Human Evolution*, *126*, 1–13.

Hammerschmidt, K., & Jürgens, U. (2007). Acoustical correlates of affective prosody. *Journal of Voice*, *21*(5), 531–540.

Hammerschmidt, W., Kulke, L., Broering, C., & Schacht, A. (2018). Money or smiles: Independent erp effects of associated monetary reward and happy faces. *PloS One*, *13*(10), e0206142.

Hammerschmidt, W., Sennhenn-Reulen, H., & Schacht, A. (2017). Associated motivational salience impacts early sensory processing of human faces. *NeuroImage*, *156*, 466–474.

Hänggi, Y. (2004). Stress and emotion recognition: An internet experiment using stress induction. *Swiss Journal of Psychology*, *63*(2), 113–125.

Harmon-Jones, E., Harmon-Jones, C., & Summerell, E. (2017). On the importance of both dimensional and discrete models of emotion. *Behavioral Sciences*, *7*(4), 66.

Hawk, S. T., Van Kleef, G. A., Fischer, A. H., & Van der Schalk, J. (2009). "worth a thousand words": Absolute and relative decoding of nonlinguistic affect vocalizations. *Emotion*, *9*(3), 293–305.

Hellbernd, N., & Sammler, D. (2016). Prosody conveys speaker's intentions: Acoustic cues for speech act perception. *Journal of Memory and Language*, *88*, 70–86.

Hess, U., Adams Jr, R., & Kleck, R. (2005). Who may frown and who should smile? dominance, affiliation, and the display of happiness and anger. *Cognition & Emotion*, *19*(4), 515–536.

Hess, U., Senécal, S., Kirouac, G., Herrera, P., Philippot, P., & Kleck, R. E. (2000). Emotional expressivity in men and women: Stereotypes and self-perceptions. *Cognition & Emotion*, *14*(5), 609–642.

Hinojosa, J., Mercado, F., & Carretié, L. (2015). N170 sensitivity to facial expression: a meta-analysis. *Neuroscience & Biobehavioral Reviews*, *55*, 498–509.

Hodges-Simeon, C. R., Gaulin, S. J., & Puts, D. A. (2010). Different vocal parameters predict perceptions of dominance and attractiveness. *Human Nature*, *21*(4), 406–427.

Hothorn, T., Hornik, K., Van De Wiel, M. A., Zeileis, A., et al. (2008). Implementing a class of permutation pests: the coin package. *Journal of Statistical Software*, *28*(8), 1–23.

Hwang, H. C., & Matsumoto, D. (2016). Measuring emotions in the face. In *Emotion measurement* (pp. 125–144). Duxford, UK: Woodhead Publishing.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, *60*(6), 581–592.

Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology*, *65*, 373–398.

Israelashvili, J., Oosterwijk, S., Sauter, D., & Fischer, A. (2019). Knowing me, knowing you: emotion differentiation in oneself is associated with recognition of others' emotions. *Cognition and Emotion*, 1–11.

Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: the negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology*, *75*(4), 887–900.

Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science*, *2*(3), 260–280.

Izard, C. E. (2010). The many meanings/aspects of emotion: Definitions, functions, activation, and regulation. *Emotion Review*, *2*(4), 363–370.

Jack, R. E., & Schyns, P. G. (2015). The human face as a dynamic tool for social communication. *Current Biology*, *25*(14), R621–R634.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning with applications in r. In G. Cassella, S. Fienberg, & I. Olkin (Eds.), *Springer texts in statistics* (pp. 303–332). New York: Springer.

Jang, D., & Elfenbein, H. A. (2015). Emotion, perception and expression of. In *International encyclopedia of the social & behavioral sciences* (2nd ed., pp. 483–489). Oxford: Elsevier.

Jessen, S., & Kotz, S. A. (2011). The temporal dynamics of processing emotions from vocal, facial, and bodily expressions. *NeuroImage*, *58*(2), 665–674.

Jessen, S., Obleser, J., & Kotz, S. A. (2012). How bodies and voices interact in early emotion perception. *PLoS One*, *7*(4), e36070.

Jiang, X., & Pell, M. D. (2014). Encoding and decoding confidence information in speech. In *Proceedings of the 7th international conference in speech prosody (social and linguistic speech prosody)* (Vol. 5762579, pp. 573–576).

Jiang, X., & Pell, M. D. (2015). On how the brain decodes vocal cues about speaker confidence. *Cortex*, *66*, 9–34.

Jiang, X., & Pell, M. D. (2017). The sound of confidence and doubt. *Speech Communication*, *88*, 106–126.

Johnstone, T., & Scherer, K. R. (2000). Vocal communication of emotion. In M. Lewis & J. Haviland (Eds.), *Handbook of emotion* (2nd ed., pp. 220–235). New York: Guildford.

Johnstone, T., Van Reekum, C. M., & Scherer, K. R. (2001). Vocal expression correlates

of appraisal processes. In *Appraisal processes in emotion: Theory, methods, research* (pp. 271–284). New York: Oxford University Press.

Jürgens, R., Fischer, J., & Schacht, A. (2018). Hot speech and exploding bombs: Autonomic arousal during emotion classification of prosodic utterances and affective sounds. *Frontiers in Psychology*, *9*, 228.

Jürgens, R., Grass, A., Drolet, M., & Fischer, J. (2015). Effect of acting experience on emotion expression and recognition in voice: Non-actors provide better stimuli than expected. *Journal of Nonverbal Behavior*, *39*(3), 195–214.

Jürgens, R., Hammerschmidt, K., & Fischer, J. (2011). Authentic and play-acted vocal emotion expressions reveal acoustic differences. *Frontiers in Psychology*, *2*, 180.

Juslin, P. N. (2013). Vocal affect expression: problems and promises. *Evolution of Emotional Communication. From sounds in nonhuman mammals to speech and music in man*, 252–273.

Juslin, P. N., & Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion*, *1*(4), 381–412.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770–814.

Juslin, P. N., & Scherer, K. R. (2005). Vocal expression of affect. In J. Harrigan, R. Rosenthal, & K. R. Scherer (Eds.), *The new handbook of methods in nonverbal behavior research* (1st ed., pp. 65–135). Oxford: Oxford University Press.

Kamboj, S. K., Krol, K. M., & Curran, H. V. (2015). A specific association between facial disgust recognition and estradiol levels in naturally cycling women. *PloS One*, *10*(4), e0122311.

Kappas, A., Hess, U., & Scherer, K. (1991). Voice and emotion. In *Fundamentals of nonverbal behavior* (pp. 200–238). Cambridge University Press.

Kelly, K. J., & Metcalfe, J. (2011). Metacognition of emotional face recognition. *Emotion*, *11*(4), 896–906.

Keltner, D., & Gross, J. J. (1999). Functional accounts of emotions. *Cognition & Emotion*, *13*(5), 467–480.

Keltner, D., & Kring, A. M. (1998). Emotion, social function, and psychopathology. *Review of General Psychology*, *2*(3), 320–342.

Keltner, D., Kring, A. M., & Bonanno, G. A. (1999). Fleeting signs of the course of life: Facial expression and personal adjustment. *Current Directions in Psychological Science*, *8*(1), 18–22.

Keltner, D., Sauter, D., Tracy, J., & Cowen, A. (2019). Emotional expression: Advances in basic emotion theory. *Journal of Nonverbal Behavior*, 1–28.

Keltner, D., Tracy, J., Sauter, D. A., Cordaro, D. C., & McNeil, G. (2016). Expression of emotion. In *Handbook of emotions* (4th ed., pp. 467–482). Guilford Press New York, NY.

Keshtiari, N., & Kuhlmann, M. (2016). The effects of culture and gender on the recognition of emotional speech: evidence from persian speakers living in a collectivist society.

International Journal of Society, Culture & Language, *4*(2), 71–86.

Kimble, C. E., & Seidel, S. D. (1991). Vocal signs of confidence. *Journal of Nonverbal Behavior*, *15*(2), 99–105.

Kitayama, S., & Ishii, K. (2002). Word and voice: Spontaneous attention to emotional speech in two cultures. *Cognition and Emotion*, *16*, 29–59.

Klasen, M., Kreifelts, B., Chen, Y.-H., Seubert, J., & Mathiak, K. (2014). Neural processing of emotion in multimodal settings. *Frontiers in Human Neuroscience*, *8*, 822.

Kohler, C. G., Walker, J. B., Martin, E. A., Healey, K. M., & Moberg, P. J. (2009). Facial emotion perception in schizophrenia: a meta-analytic review. *Schizophrenia Bulletin*, *36*(5), 1009–1019.

Kohler, K. J. (2008). 'speech-smile','speech-laugh','laughter'and their sequencing in dialogic interaction. *Phonetica*, *65*(1-2), 1–18.

Kordsmeyer, T. L., Lohöfener, M., & Penke, L. (2019). Male facial attractiveness, dominance, and health and the interaction between cortisol and testosterone. *Adaptive Human Behavior and Physiology*, *5*(1), 1–12.

Koriat, A. (2008). When confidence in a choice is independent of which choice is made. *Psychonomic Bulletin & Review*, *15*(5), 997–1001.

Koriat, A., & Levy-Sadot, R. (1999). Processes underlying metacognitive judgments: Information-based and experience-based monitoring of one's own knowledge. In *Dual-process theories in social psychology* (pp. 483–502). New York: Guilford Press.

Kosonogov, V., & Titova, A. (2018). Recognition of all basic emotions varies in accuracy and reaction time: A new verbal method of measurement. *International Journal of Psychology*(16).

Kotz, S. A., & Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: Erp evidence. *Brain Research*, *1151*, 107–118.

Kraus, M. W. (2017). Voice-only communication enhances empathic accuracy. *American Psychologist*, *72*(7), 644–654.

Kraus, M. W., Park, J. W., & Tan, J. J. (2017). Signs of social class: The experience of economic inequality in everyday life. *Perspectives on Psychological Science*, *12*(3), 422–435.

Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., & Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: an event-related fmri study. *NeuroImage*, *37*(4), 1445–1456.

Kret, M. E., & De Gelder, B. (2012). A review on sex differences in processing emotional signals. *Neuropsychologia*, *50*(7), 1211–1221.

Krumhuber, E. G., Kappas, A., & Manstead, A. S. (2013). Effects of dynamic aspects of facial expressions: A review. *Emotion Review*, *5*(1), 41–46.

Krumhuber, E. G., & Scherer, K. R. (2011). Affect bursts: dynamic patterns of facial expression. *Emotion*, *11*(4), 825–841.

Kuppens, P., Tuerlinckx, F., Russell, J. A., & Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological Bulletin*, *139*(4), 917–

940.

LaFrance, M., & Banaji, M. (1992). Toward a reconsideration of the gender-emotion relationship. *Emotion and social behavior*, *14*, 178–201.

Lambrecht, L., Kreifelts, B., & Wildgruber, D. (2014). Gender differences in emotion recognition: Impact of sensory modality and emotional category. *Cognition & Emotion*, *28*(3), 452–469.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & Van Knippenberg, A. (2010). Presentation and validation of the radboud faces database. *Cognition and Emotion*, *24*(8), 1377–1388.

Latinus, M., & Belin, P. (2011). Human voice perception. *Current Biology*, *21*(4), R143–R145.

Laukka, P. (2005). Categorical perception of vocal emotion expressions. *Emotion*, *5*(3), 277–295.

Laukka, P. (2008). Research on vocal expression of emotion: State of the art and future directions. In *Emotions in the human voice* (pp. 153–171). Abingdon, UK: Plural Publishing.

Laukka, P., Elfenbein, H. A., Thingujam, N. S., Rockstuhl, T., Iraki, F. K., Chui, W., & Althoff, J. (2016). The expression and recognition of emotions in the voice across five nations: A lens model analysis based on acoustic features. *Journal of Personality and Social Psychology*, *111*(5), 686–705.

Lausen, A., Hammerschmidt, K., & Schacht, A. (2019). Emotion recognition and confidence ratings predicted by vocal stimulus type and acoustic parameters. *PsyArXiv*.

Lausen, A., & Schacht, A. (2018). Gender differences in the recognition of vocal emotions. *Frontiers in Psychology*, *9*, 882.

Lee, N. C., Krabbendam, L., White, T. P., Meeter, M., Banaschewski, T., Barker, G. J., ... others (2013). Do you see what i see? sex differences in the discrimination of facial emotions during adolescence. *Emotion*, *13*(6), 1030–1040.

Levenson, R. W. (2011). Basic emotion questions. *Emotion Review*, *3*(4), 379–386.

Levenson, R. W., & Ruef, A. M. (1992). Empathy: A physiological substrate. *Journal of Personality and Social Psychology*, *63*(2), 234–246.

Lima, C. F., Alves, T., Scott, S. K., & Castro, S. L. (2014). In the ear of the beholder: How age shapes emotion processing in nonverbal vocalizations. *Emotion*, *14*(1), 145–160.

Lindquist, K. A. (2013). Emotions emerge from more basic psychological ingredients: A modern psychological constructionist model. *Emotion Review*, *5*(4), 356–368.

Liu, T., Pinheiro, A. P., Deng, G., Nestor, P. G., McCarley, R. W., & Niznikiewicz, M. A. (2012). Electrophysiological insights into processing nonverbal emotional vocalizations. *NeuroReport*, *23*(2), 108–112.

Lyusin, D., & Ovsyannikova, V. (2016). Measuring two aspects of emotion recognition ability: Accuracy vs. sensitivity. *Learning and Individual Differences*, *52*, 129–136.

Marsh, A. A., & Blair, R. J. R. (2008). Deficits in facial affect recognition among antisocial populations: a meta-analysis. *Neuroscience & Biobehavioral Reviews*, *32*(3), 454–465.

Martinez, A., & Du, S. (2012). A model of the perception of facial expressions of emotion by humans: Research overview and perspectives. *Journal of Machine Learning Research*, *13*(May), 1589–1608.

Massaro, D. W., & Egan, P. B. (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin & Review*, *3*(2), 215–221.

Matsumoto, D., & Hwang, H. S. (2011). Judgments of facial expressions of emotion in profile. *Emotion*, *11*(5), 1223–1229.

Matsumoto, D., & Lee, M. (1993). Consciousness, volition, and the neuropsychology of facial expressions of emotion. *Consciousness and Cognition: An International Journal*, *2*(3), 237–254.

Matsumoto, D., LeRoux, J., Wilson-Cohn, C., Raroque, J., Kooken, K., Ekman, P., ... others (2000). A new test to measure emotion recognition ability: Matsumoto and ekman's japanese and caucasian brief affect recognition test (jacbart). *Journal of Nonverbal Behavior*, *24*(3), 179–209.

Matt, G. E., Vázquez, C., & Campbell, W. K. (1992). Mood-congruent recall of affectively toned stimuli: A meta-analytic review. *Clinical Psychology Review*, *12*(2), 227–255.

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? what does "failure to replicate" really mean? *American Psychologist*, *70*(6), 487.

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2001). Emotional intelligence as a standard intelligence. *Emotion*, *1*, 232–242.

Mazur, A., & Booth, A. (2014). Testosterone is related to deviance in male army veterans, but relationships are not moderated by cortisol. *Biological Psychology*, *96*, 72–76.

McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin*, *126*(3), 424–453.

McDuff, D., Kodra, E., el Kaliouby, R., & LaFrance, M. (2017). A large-scale analysis of sex differences in facial expressions. *PloS One*, *12*(4), e0173942.

Mehta, P. H., & Josephs, R. A. (2010). Testosterone and cortisol jointly regulate dominance: Evidence for a dual-hormone hypothesis. *Hormones and Behavior*, *58*(5), 898–906.

Mehta, P. H., & Prasad, S. (2015). The dual-hormone hypothesis: a brief review and future research agenda. *Current Opinion in Behavioral Sciences*, *3*, 163–168.

Mehta, P. H., Wuehrmann, E. V., & Josephs, R. A. (2009). When are low testosterone levels advantageous? the moderating role of individual versus intergroup competition. *Hormones and Behavior*, *56*(1), 158–162.

Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(4), 851–861.

Mitchell, R. L., Elliott, R., Barry, M., Cruttenden, A., & Woodruff, P. W. (2003). The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia*, *41*(10), 1410–1421.

Mitchell, R. L., & Phillips, L. H. (2015). The overlapping relationship between emotion perception and theory of mind. *Neuropsychologia*, *70*, 1–10.

Montagne, B., Kessels, R. P., De Haan, E. H., & Perrett, D. I. (2007). The emotion recognition task: A paradigm to measure the perception of facial emotional expressions at different intensities. *Perceptual and Motor Skills*, *104*(2), 589–598.

Montagne, B., Kessels, R. P., Frigerio, E., de Haan, E. H., & Perrett, D. I. (2005). Sex differences in the perception of affective facial expressions: Do men really lack emotional sensitivity? *Cognitive processing*, *6*(2), 136–141.

Morningstar, M. (2017). *Age-related differences in the production and recognition of vocal socio-emotional expressions* (Unpublished doctoral dissertation). McGill University, Montreal.

Mortillaro, M., Meuleman, B., & Scherer, K. R. (2012). Advocating a componential appraisal model to guide emotion recognition. *International Journal of Synthetic Emotions*, *3*(1), 18–32.

Mozziconacci, S. (2002). Prosody and emotions. In *Proceedings of speech prosody* (pp. 1–9).

Müri, R. M. (2016). Cortical control of facial expression. *Journal of Comparative Neurology*, *524*(8), 1578–1585.

Murphy, F. C., Nimmo-Smith, I., & Lawrence, A. D. (2003). Functional neuroanatomy of emotions: a meta-analysis. *Cognitive, Affective, & Behavioral Neuroscience*, *3*(3), 207–233.

Murray, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America*, *93*(2), 1097–1108.

Noroozi, F., Sapiński, T., Kamińska, D., & Anbarjafari, G. (2017). Vocal-based emotion recognition using random forests and decision tree. *International Journal of Speech Technology*, *20*(2), 239–246.

Nummenmaa, L., & Calvo, M. G. (2015). Dissociation between recognition and detection advantage for facial expressions: A meta-analysis. *Emotion*, *15*(2), 243–256.

Nygaard, L. C., & Lunders, E. R. (2002). Resolution of lexical ambiguity by emotional tone of voice. *Memory & Cognition*, *30*(4), 583–593.

Öhman, A. (1993). Fear and anxiety as emotional phenomena: Clinical phenomenology, evolutionary perspectives, and information-processing mechanisms. *Handbook of Emotions*.

Olderbak, S., & Wilhelm, O. (2017). Emotion perception and empathy: An individual differences test of relations. *Emotion*, *17*(7), 1092–1106.

Oveis, C., Spectre, A., Smith, P. K., Liu, M. Y., & Keltner, D. (2016). Laughter conveys status. *Journal of Experimental Social Psychology*, *65*, 109–115.

Owren, M. J. (2008). Gsu praat tools: Scripts for modifying and analyzing sounds using praat acoustics software. *Behavior Research Methods*, *40*(3), 822–829.

Palermo, R., & Coltheart, M. (2004). Photographs of facial expression: Accuracy, response times, and ratings of intensity. *Behavior Research Methods, Instruments, &*

Computers, *36*(4), 634–638.

Panksepp, J. (2007). Neuroevolutionary sources of laughter and social joy: Modeling primal human laughter in laboratory rats. *Behavioural Brain Research*, *182*(2), 231–244.

Parkins, R. (2012). Gender and emotional expressiveness: An analysis of prosodic features in emotional expression. *Pragmatics and Intercultural Communication*, *5*(1), 46–54.

Parr, L. A. (2003). The discrimination of faces and their emotional content by chimpanzees (pan troglodytes). *Annals of the New York Academy of Sciences*, *1000*(1), 56–78.

Parr, L. A., Cohen, M., & De Waal, F. (2005). Influence of social context on the use of blended and graded facial displays in chimpanzees. *International Journal of Primatology*, *26*(1), 73–103.

Parr, L. A., Waller, B. M., & Fugate, J. (2005). Emotional communication in primates: implications for neurobiology. *Current Opinion in Neurobiology*, *15*(6), 716–720.

Parsons, C. E., Young, K. S., Craske, M. G., Stein, A. L., & Kringelbach, M. L. (2014). Introducing the oxford vocal (oxvoc) sounds database: a validated set of non-acted affective sounds from human infants, adults, and domestic animals. *Frontiers in Psychology*, *5*, 562.

Parsons, C. E., Young, K. S., Joensson, M., Brattico, E., Hyam, J. A., Stein, A., ... Kringelbach, M. L. (2014). Ready for action: a role for the human midbrain in responding to infant vocalizations. *Social Cognitive and Affective Neuroscience*, *9*(7), 977–984.

Patel, S., Scherer, K. R., Björkner, E., & Sundberg, J. (2011). Mapping emotions into acoustic space: The role of voice production. *Biological Psychology*, *87*(1), 93–98.

Paulmann, S. (2016). The neurocognition of prosody. In G. Hickok & S. Small (Eds.), *Neurobiology of language* (pp. 1109–1120). San Diego: Elsevier.

Paulmann, S., Furnes, D., Bøkenes, A. M., & Cozzolino, P. J. (2016). How psychological stress affects emotional prosody. *PloS One*, *11*(11), e0165022.

Paulmann, S., & Kotz, S. A. (2008). An erp investigation on the temporal dynamics of emotional prosody and emotional semantics in pseudo-and lexical-sentence context. *Brain and Language*, *105*(1), 59–69.

Paulmann, S., & Pell, M. D. (2010). Dynamic emotion processing in parkinson's disease as a function of channel availability. *Journal of Clinical and Experimental Neuropsychology*, *32*(8), 822–835.

Paulmann, S., & Pell, M. D. (2011). Is there an advantage for recognizing multi-modal emotional stimuli? *Motivation and Emotion*, *35*(2), 192–201.

Paulmann, S., Pell, M. D., & Kotz, S. A. (2008). How aging affects the recognition of emotional speech. *Brain and Language*, *104*(3), 262–269.

Paulmann, S., Titone, D., & Pell, M. D. (2012). How emotional prosody guides your way: evidence from eye movements. *Speech Communication*, *54*(1), 92–107.

Paulmann, S., & Uskul, A. K. (2014). Cross-cultural emotional prosody recognition: Evidence from chinese and british listeners. *Cognition & Emotion*, *28*(2), 230–244.

Peeters, G., & Czapinski, J. (1990). Positive-negative asymmetry in evaluations: The distinction between affective and informational negativity effects. In W. Stroebe

& M. Hewstone (Eds.), *European review of social psychology* (Vol. 1, pp. 33–60). Chichester, England: Wiley.

Pell, M. D. (2002). Evaluation of nonverbal emotion in face and voice: Some preliminary findings on a new battery of tests. *Brain and Cognition*, *48*(2-3), 499–504.

Pell, M. D., & Kotz, S. A. (2011). On the time course of vocal emotion recognition. *PLoS One*, *6*(11), e27256.

Pell, M. D., Kotz, S. A., Paulmann, S., & Alasseri, A. (2005). Recognition of basic emotions from speech prosody as a function of language and sex. In *Psychonomic society 46th annual meeting* (Vol. 10, pp. 97–98).

Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior*, *33*(2), 107–120.

Pell, M. D., Paulmann, S., Dara, C., Alasseri, A., & Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics*, *37*(4), 417–435.

Pell, M. D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., & Rigoulot, S. (2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biological Psychology*, *111*, 14–25.

Pichon, S., de Gelder, B., & Grezes, J. (2008). Emotional modulation of visual and motor areas by dynamic body expressions of anger. *Social Neuroscience*, *3*(3-4), 199–212.

Pichora-Fuller, M. K., Dupuis, K., & Van Lieshout, P. (2016). Importance of f0 for predicting vocal emotion categorization. *The Journal of the Acoustical Society of America*, *140*(4), 3401–3401.

Piwek, L., Pollick, F., & Petrini, K. (2015). Audiovisual integration of emotional signals from others' social interactions. *Frontiers in Psychology*, *6*, 611.

Planalp, S. (1999). *Communicating emotion: Social, moral, and cultural processes*. Cambridge University Press.

Plant, E. A., Hyde, J. S., Keltner, D., & Devine, P. G. (2000). The gender stereotyping of emotions. *Psychology of Women Quarterly*, *24*(1), 81–92.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903.

Ponsot, E., Burred, J. J., Belin, P., & Aucouturier, J.-J. (2018). Cracking the social code of speech prosody using reverse correlation. *Proceedings of the National Academy of Sciences*, *115*(15), 3972–3977.

Posamentier, M. T., & Abdi, H. (2003). Processing faces and facial expressions. *Neuropsychology Review*, *13*(3), 113–143.

Rahman, Q., Wilson, G. D., & Abrahams, S. (2004). Sex, sexual orientation, and identification of positive and negative facial affect. *Brain and Cognition*, *54*(3), 179–185.

Raithel, V., & Hielscher-Fastabend, M. (2004). Emotional and linguistic perception of prosody. reception of prosody. *Folia Phoniatrica et Logopaedica*, *56*(1), 7–13.

R Core Team, R. (2017). *R: A language and environment for statistical computing. r foundation for statistical computing*. Vienna, Austria. 2017.

Recio, G., Schacht, A., & Sommer, W. (2014). Recognizing dynamic facial expressions of emotion: Specificity and intensity effects in event-related brain potentials. *Biological Psychology*, *96*, 111–125.

Recio, G., Sommer, W., & Schacht, A. (2011). Electrophysiological correlates of perceiving and evaluating static and dynamic facial emotional expressions. *Brain Research*, *1376*, 66–75.

Richard, F. D., Bond Jr, C. F., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, *7*(4), 331–363.

Rigoulot, S., Wassiliwizky, E., & Pell, M. D. (2013). Feeling backwards? how temporal order in speech affects the time course of vocal emotion recognition. *Frontiers in Psychology*, *4*, 367.

Riviello, M. T., & Esposito, A. (2016). Recognition performance on the american and italian cross-modal databases. In M. T. Riviello & A. Esposito (Eds.), *On the perception of dynamic emotional expressions: A cross-cultural comparison* (pp. 9–23). Springer.

Roberts, R. D., MacCann, C., Matthews, G., & Zeidner, M. (2010). Emotional intelligence: Toward a consensus of models and measures. *Social and Personality Psychology Compass*, *4*(10), 821–840.

Rosenthal, R., & DePaulo, B. M. (1979). Sex differences in accommodation in nonverbal communication. In R. Rosenthal (Ed.), *Skill in nonverbal communication: Individual differences* (pp. 68–96). Oelgeschlager, Gunn & Hain Publishers.

Rotter, N. G., & Rotter, G. S. (1988). Sex differences in the encoding and decoding of negative facial emotions. *Journal of Nonverbal Behavior*, *12*(2), 139–148.

Rukavina, S., Sachsenweger, F., Jerg-Bretzke, L., Daucher, A. E., Traue, H. C., Walter, S., & Hoffmann, H. (2018). Testosterone and its influence on emotion recognition in young, healthy males. *Psychology*, *9*(07), 1814–1827.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? a review of the cross-cultural studies. *Psychological Bulletin*, *115*(1), 102–141.

Sarkar, A., Mehta, P. H., & Josephs, R. A. (2019). The dual-hormone approach to dominance and status-seeking. In O. C. Schultheiss & P. H. Mehta (Eds.), *Routledge international handbook of social neuroendocrinology* (1st ed., pp. 113–132). Abingdon, UK: Routledge.

Sauter, D. A. (2006). *An investigation into vocal expressions of emotions: the roles of valence, culture, and acoustic factors.* (Unpublished doctoral dissertation). University College London.

Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *Quarterly Journal of Experimental Psychology*, *63*(11), 2251–2272.

Sauter, D. A., & Fischer, A. H. (2018). Can perceivers recognise emotions from spontaneous expressions? *Cognition & Emotion*, *32*(3), 504–515.

Sauter, D. A., Panattoni, C., & Happé, F. (2013). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology*, *31*(1), 97–113.

Sbattella, L., Colombo, L., Rinaldi, C., Tedesco, R., Matteucci, M., & Trivilini, A. (2014). Extracting emotions and communication styles from prosody. In H. Da Silva, A. Holzinger, S. Fairclough, & D. Majoe (Eds.), *Physiological computing systems, 8908* (pp. 21–42).

Scarantino, A., & de Sousa, R. (2018). Emotion. In *The stanford encyclopedia of philosophy.*

Schacht, A., & Sommer, W. (2009a). Emotions in word and face processing: early and late cortical responses. *Brain and Cognition*, *69*(3), 538–550.

Schacht, A., & Sommer, W. (2009b). Time course and task dependence of emotion effects in word processing. *Cognitive, Affective, & Behavioral Neuroscience*, *9*(1), 28–43.

Schaerlaeken, S., & Grandjean, D. (2018). Unfolding and dynamics of affect bursts decoding in humans. *PloS One*, *13*(10), e0206216.

Scherer, K. (2009). Emotion definitions (psychological perspectives). In *The oxford companion to emotion and the affective sciences* (1st ed., pp. 145–149). Oxford University Press.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, *99*(2), 143–165.

Scherer, K. R. (1989). Vocal correlates of emotional arousal and affective disturbance. In *Handbook of social psychophysiology* (pp. 165–197). John Wiley & Sons.

Scherer, K. R. (1994). Affect bursts. In S. H. Van Goozen, N. E. Van de Poll, & J. A. Sergeant (Eds.), *Emotions: Essays on emotion theory* (pp. 161–193). Hillsdale, NJ: Erlbaum.

Scherer, K. R. (2009). Emotions are emergent processes: they require a dynamic computational architecture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3459–3474.

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, *32*(1), 76–92.

Scherer, K. R., & Bänziger, T. (2004). Emotional expression in prosody: a review and an agenda for future research. In *Speech prosody 2004, international conference.*

Scherer, K. R., Clark-Polner, E., & Mortillaro, M. (2011). In the eye of the beholder? universality and cultural specificity in the expression and perception of emotion. *International Journal of Psychology*, *46*(6), 401–435.

Scherer, K. R., & Ellgring, H. (2007). Multimodal expression of emotion: Affect programs or componential appraisal patterns? *Emotion*, *7*(1), 158–171.

Scherer, K. R., London, H., & Wolf, J. J. (1973). The voice of confidence: Paralinguistic cues and audience evaluation. *Journal of Research in Personality*, *7*(1), 31–44.

Scherer, K. R., & Moors, A. (2019). The emotion process: Event appraisal and component differentiation. *Annual Review of Psychology*, *70*, 719–745.

Scherer, K. R., et al. (2000). Psychological models of emotion. In *The neuropsychology of*

emotion (pp. 137–162). Oxford University Press.

Scherer, K. R., & Scherer, U. (2011). Assessing the ability to recognize facial and vocal expressions of emotion: Construction and validation of the emotion recognition index. *Journal of Nonverbal Behavior*, *35*(4), 305–326.

Schirmer, A. (2013). Sex differences in emotion. *The Cambridge handbook of human affective neuroscience*, 591–610.

Schirmer, A., & Adolphs, R. (2017). Emotion perception from face, voice, and touch: comparisons and convergence. *Trends in Cognitive Sciences*, *21*(3), 216–228.

Schirmer, A., & Kotz, S. A. (2003). Erp evidence for a sex-specific stroop effect in emotional speech. *Journal of Cognitive Neuroscience*, *15*(8), 1135–1148.

Schlegel, K., Fontaine, J. R., & Scherer, K. R. (2017). The nomological network of emotion recognition ability: Evidence from the geneva emotion recognition test. *European Journal of Psychological Assessment*, 1–12.

Schmid, P. C., & Mast, M. S. (2010). Mood effects on emotion recognition. *Motivation and Emotion*, *34*(3), 288–292.

Schmidt, F. L., & Oh, I.-S. (2013). Methods for second order meta-analysis and illustrative applications. *Organizational Behavior and Human Decision Processes*, *121*(2), 204–218.

Schmidt, K. L., & Cohn, J. F. (2001). Human facial expressions as adaptations: Evolutionary questions in facial expression research. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, *116*(S33), 3–24.

Schultheiss, O. C., Dlugash, G., & Mehta, P. H. (2019). Hormone measurement in social neuroendocrinology: a comparison of immunoassay and mass spectrometry methods. In O. C. Schultheiss & P. H. Mehta (Eds.), *Routledge international handbook of social neuroendocrinology* (1st ed., pp. 26–41). Abingdon, UK: Routledge.

Schultheiss, O. C., Schiepe, A., & Rawolle, M. (2012). Hormone assays. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *Handbook of research methods in psychology* (1st ed., pp. 489–500). Washington D.C.: American Psychological Association.

Schultheiss, O. C., & Stanton, S. J. (2009). Assessment of salivary hormones. In E. Harmon-Jones & J. S. Beer (Eds.), *Methods in social neuroscience* (pp. 17–44). New York: Guilford Press.

Schwartz, R., & Pell, M. D. (2012). Emotional speech processing at the intersection of prosody and semantics. *PloS One*, *7*(10), e47279.

Scott, S. K., Sauter, D., & McGettigan, C. (2010). Brain mechanisms for processing perceived emotional vocalizations in humans. In *Handbook of behavioral neuroscience* (Vol. 19, pp. 187–197). Elsevier.

Seifritz, E., Esposito, F., Neuhoff, J. G., Lüthi, A., Mustovic, H., Dammann, G., ... others (2003). Differential sex-independent amygdala response to infant crying and laughing in parents versus nonparents. *Biological Psychiatry*, *54*(12), 1367–1375.

Shariff, A. F., & Tracy, J. L. (2011a). Emotion expressions: On signals, symbols, and

spandrels—a response to barrett (2011). *Current Directions in Psychological Science*, *20*(6), 407–408.

Shariff, A. F., & Tracy, J. L. (2011b). What are emotion expressions for? *Current Directions in Psychological Science*, *20*(6), 395–399.

Shaver, P., Schwartz, J., Kirson, D., & O'connor, C. (1987). Emotion knowledge: further exploration of a prototype approach. *Journal of Personality and Social Psychology*, *52*(6), 1061–1086.

Shields, S. A. (2002). *Speaking from the heart: Gender and the social meaning of emotion.* Cambridge University Press.

Shih, C., & Kochanski, G. (2002). Icslp2002 prosody and prosodic modeling. three-hour tutorial at the seventh international conference on spoken language processing. *Denver CO, 9/15/2002*.

Shuman, V., & Scherer, K. R. (2014). Concepts and structures of emotions. In *International handbook of emotions in education* (pp. 13–35). New York: Routledge/Taylor & Francis Group.

Shuman, V., & Scherer, K. R. (2015). Psychological structure of emotions. In *International encyclopedia of the social and behavioral sciences* (2nd ed., Vol. 7, pp. 526–533). Oxford: Elsevier.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366.

Simon-Thomas, E., Sauter, D., Sinicropi-Yao, L., Abramson, A., & Keltner, D. (2007). Vocal bursts communicate discrete emotions: Evidence for new displays. *Nature Precedings*.

Smith, F. W., & Rossit, S. (2018). Identifying and detecting facial expressions of emotion in peripheral vision. *PloS One*, *13*(5), e0197160.

Smoski, M., & Bachorowski, J.-A. (2003). Antiphonal laughter between friends and strangers. *Cognition and Emotion*, *17*(2), 327–340.

Snowdon, C. T. (2003). Expression of emotion in nonhuman animals. In *Handbook of affective sciences* (pp. 457–480). New York: Oxford University Press.

Sokolov, A. A., Krüger, S., Enck, P., Krägeloh-Mann, I., & Pavlova, M. A. (2011). Gender affects body language reading. *Frontiers in Psychology*, *2*, 16.

Sokolowski, K., Schmalt, H.-D., Langens, T. A., & Puca, R. M. (2000). Assessing achievement, affiliation, and power motives all at once: The multi-motive grid (mmg). *Journal of Personality Assessment*, *74*(1), 126–145.

Šolcová, I. P., & Lačev, A. (2017). Differences in male and female subjective experience and physiological reactions to emotional stimuli. *International Journal of Psychophysiology*, *117*, 75–82.

Soto, J. A., & Levenson, R. W. (2009). Emotion recognition across cultures: The influence of ethnicity on empathic accuracy and physiological linkage. *Emotion*, *9*(6), 874–884.

Tackett, J. L., Reardon, K. W., Herzhoff, K., Page-Gould, E., Harden, K. P., & Josephs, R. A. (2015). Estradiol and cortisol interactions in youth externalizing psychopathol-

ogy. *Psychoneuroendocrinology*, *55*, 146–153.

Teixeira, J. P., & Gonçalves, A. (2014). Accuracy of jitter and shimmer measurements. *Procedia Technology*, *16*, 1190–1199.

Terracciano, A., Merritt, M., Zonderman, A. B., & Evans, M. K. (2003). Personality traits and sex differences in emotion recognition among african americans and caucasians. *Annals of the New York Academy of Sciences*, *1000*(1), 309–312.

Thompson, A. E., & Voyer, D. (2014). Sex differences in the ability to recognise non-verbal displays of emotion: A meta-analysis. *Cognition & Emotion*, *28*(7), 1164–1195.

Thompson, W. F., & Balkwill, L.-L. (2006). Decoding speech prosody in five languages. *Semiotica*, *2006*(158), 407–424.

Thompson, W. F., & Balkwill, L. L. (2009). Cross-cultural similarities and differences. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 755–791). New York: Oxford University Press.

Timmers, M., Fischer, A., & Manstead, A. (2003). Ability versus vulnerability: Beliefs about men's and women's emotional behaviour. *Cognition & Emotion*, *17*(1), 41–63.

Toivanen, J., Väyrynen, E., & Seppänen, T. (2004). Automatic discrimination of emotion from spoken finnish. *Language and Speech*, *47*(4), 383–412.

Toivanen, J., Väyrynen, E., & Seppänen, T. (2005). Gender differences in the ability to discriminate emotional content from speech. In *Proc. fonetik* (pp. 119–122).

Tooby, J., & Cosmides, L. (2008). The evolutionary psychology of the emotions and their relationship to internal regulatory variables. In *Handbook of emotions* (3rd ed., pp. 114–137). New York: Guilford Press.

Van den Stock, J., Righart, R., & De Gelder, B. (2007). Body expressions influence recognition of emotions in the face and voice. *Emotion*, *7*(3), 487–494.

Van Honk, J., & JLG Schutter, D. (2007). Testosterone reduces conscious detection of signals serving social correction: Implications for antisocial behavior. *Psychological Science*, *18*(8), 663–667.

Van Kleef, G. (2016). Emotions as social information theory. In *The interpersonal dynamics of emotion: Toward an integrative theory of emotions as social information* (pp. 11–78). Cambridge: Cambridge University Press.

Van Kleef, G. (2017). The social effects of emotions are functionally equivalent across expressive modalities. *Psychological Inquiry*, *28*(2-3), 211–216.

Viau, V. (2002). Functional cross-talk between the hypothalamic-pituitary-gonadal and-adrenal axes. *Journal of Neuroendocrinology*, *14*(6), 506–513.

Võ, M. L., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The berlin affective word list reloaded (bawl-r). *Behavior Research Methods*, *41*(2), 534–538.

Vongas, J. G., & Al Hajj, R. (2017). The effects of competition and implicit power motive on men's testosterone, emotion recognition, and aggression. *Hormones and Behavior*, *92*, 57–71.

Vroomen, J., Driver, J., & De Gelder, B. (2001). Is cross-modal integration of emotional expressions independent of attentional resources? *Cognitive, Affective, & Behavioral*

Neuroscience, *1*(4), 382–387.

Vytal, K., & Hamann, S. (2010). Neuroimaging support for discrete neural correlates of basic emotions: a voxel-based meta-analysis. *Journal of Cognitive Neuroscience*, *22*(12), 2864–2885.

Waaramaa, T. (2017). Gender differences in identifying emotions from auditory and visual stimuli. *Logopedics Phoniatrics Vocology*, *42*(4), 160–166.

Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior*, *17*(1), 3–28.

Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, *27*(3), 236–247.

Welker, K. M., Lozoya, E., Campbell, J. A., Neumann, C. S., & Carré, J. M. (2014). Testosterone, cortisol, and psychopathic traits in men and women. *Physiology & Behavior*, *129*, 230–236.

Wells, L. J., Gillespie, S. M., & Rotshtein, P. (2016). Identification of emotional facial expressions: effects of expression, intensity, and sex on eye gaze. *PloS One*, *11*(12), e0168307.

Williams, L. M., Mathersul, D., Palmer, D. M., Gur, R. C., Gur, R. E., & Gordon, E. (2009). Explicit identification and implicit recognition of facial emotions: I. age effects in males and females across 10 decades. *Journal of Clinical and Experimental Neuropsychology*, *31*(3), 257–277.

Williams, M. A., & Mattingley, J. B. (2006). Do angry men get noticed? *Current Biology*, *16*(11), R402–R404.

Wilson, T. D., & Gilbert, D. T. (2008). Explaining away: A model of affective adaptation. *Perspectives on Psychological Science*, *3*(5), 370–386.

Wingenbach, T. S., Ashwin, C., & Brosnan, M. (2018). Sex differences in facial emotion recognition across varying expression intensity levels from videos. *PLoS One*, *13*(1), e0190634.

Wyczesany, M., & Ligeza, T. S. (2015). Towards a constructionist approach to emotions: verification of the three-dimensional model of affect with eeg-independent component analysis. *Experimental Brain Research*, *233*(3), 723–733.

Xu, Y., Lee, A., Wu, W.-L., Liu, X., & Birkholz, P. (2013). Human vocal attractiveness as signaled by body size projection. *PloS One*, *8*(4), e62397.

Young, K. S., Parsons, C. E., LeBeau, R. T., Tabak, B. A., Sewart, A. R., Stein, A., ... Craske, M. G. (2017). Sensing emotion in voices: Negativity bias and gender differences in a validation study of the oxford vocal ('oxvoc') sounds database. *Psychological Assessment*, *29*(8), 967–977.

Zaki, J., Bolger, N., & Ochsner, K. (2008). It takes two: The interpersonal nature of empathic accuracy. *Psychological Science*, *19*(4), 399–404.

Zell, E., & Krizan, Z. (2014). Do people have insight into their abilities? a metasynthesis. *Perspectives on Psychological Science*, *9*(2), 111–125.

Zuckerman, M., Lipets, M. S., Koivumaki, J. H., & Rosenthal, R. (1975). Encoding and

decoding nonverbal cues of emotion. *Journal of Personality and Social Psychology*, *32*(6), 1068–1076.

Zupan, B., Babbage, D., Neumann, D., & Willer, B. (2017). Sex differences in emotion recognition and emotional inferencing following severe traumatic brain injury. *Brain Impairment*, *18*(1), 36–48.

APPENDIX: STUDY 1

Paulmann Pseudo-sentences

Table A1.1: Pseudo-Sentences | *Relative Frequencies* (RF), *Standard Error* (SE), *95% Confidence Intervals* (CI) for all **pseudo-sentences** & emotions ($N = 145$)

Pseudo-sentences	*Emotions*	*RF*	*SE*	*CI95%*
Hung set das Vermalet gereubt ind verprusst. (a06)		.890	.018	[.854, .926]
Hung set die Noschichte geballigt ind geschweugen. (a13)		.728	.026	[.676, .779]
Hung set das Antirgen verbirken ind ramgemuckert. (a20)		.890	.018	[.854, .926]
Hung set die Jigundlachen beligen ind nagebrucht. (a26)		.866	.020	[.826, .905]
Mon set das Portan nogebrennt ind wiggewarfen. (a29)	*Angry*	.817	.023	[.773, .862]
Hung set die Ultkraft getreunigt ind keuniiert. (a31)		.814	.023	[.769, .859]
Mon set die Helbtarges nogeläßt ind nogezuttelt. (a44)		.817	.023	[.773, .862]
Mon set die Hiest gelabbt ind gekreun. (a46)		.890	.018	[.854, .926]
Mon set den Urzäk getruten ind vergeien. (a48)		.969	.010	[.949, .989]
Hung set die Willo bewöcht ind verkeustet. (a49)		.666	.028	[.611, .720]
Hung set die Spulza verbrutet ind nogelackt. (d01)		.500	.029	[.442, .558]
Hung set die Millhulde bewehnt ind gepfunken. (d03)		.434	.029	[.377, .492]
Mon set den Luchdel nogegruben ind unspizart. (d06)		.255	.026	[.205, .305]
Hung set die Liche gezäckt ind ubgedackt. (d08)		.403	.029	[.347, .460]
Hung set die Busche geweiden ind gepfunken. (d19)	*Disgust*	.548	.029	[.491, .606]
Hung set die Uls getutschelt ind matgenimmen. (d27)		.517	.029	[.460, .575]
Mon set die Titun terstiert ind eungepuckt. (d30)		.834	.022	[.792, .877]
Hung set die Quadrul verrinlussigt ind gepfunken. (d35)		.441	.029	[.384, .499]
Mon set das Pust izbrichen ind unspizart. (d42)		.490	.029	[.432, .547]
Mon set das Edelsteist geschlubbt ind unspizart. (d46)		.303	.027	[.251, .356]
Hung set den Wiffecke bejaubt ind nogegraffen. (f04)		.471	.029	[.413, .528]
Mon set das Bakobi gedellen ind gezagen. (f05)		.790	.024	[.743, .837]
Hung set die Zamiat gewaungt ind ubgewarfen. (f06)		.545	.029	[.488, .602]
Hung set die Schimme vermännen ind geurdet. (f14)	*Fear*	.814	.023	[.769, .859]
Mon set den Kneiloparden geknüfft ind verschruckt. (f17)		.476	.029	[.418, .533]
Hung set die Batschoft bedreht ind ingezöndet. (f22)		.446	.029	[.389, .504]
Mon set die Ralle benatzt ind wiggetiermt. (f25)		.459	.029	[.401, .516]
Hung set den Altistark beteiglet ind gezattert. (f35)		.731	.026	[.680, .782]
Hung set den Rackwig verknirrt ind ubgedankelt. (f46)		.503	.029	[.446, .561]
Hung set die Nate geklettet ind izprasst. (f47)		.431	.029	[.374, .488]
Hung set die Welstbare geruft ind ingefegt. (h04)		.528	.029	[.470, .585]
Hung set den Harindisan belonkt ind geheichelt. (h07)		.566	.029	[.508, .623]
Hung set das Puchel izkunnt ind gekobelt. (h09)		.472	.029	[.415, .530]
Hung set die Gahli döllerwultigt ind geheichelt. (h10)		.628	.028	[.572, .683]
Hung set den Fiebele geteiret ind geflötzt. (h15)	*Happy*	.769	.025	[.720, .817]
Mon set den Akosent gebösten ind gepförmt. (h17)		.466	.029	[.408, .523]
Hung set das Bil geberent ind geheichelt. (h22)		.783	.024	[.735, .830]
Hung set den Pürer getöscht ind gekobelt. (h26)		.807	.023	[.761, .852]
Mon set den Schindt geheuritet ind gepförmt.(h30)		.538	.029	[.481, .595]
Hung set das Einsart bewindert ind fatagrofiert (h39)		.459	.029	[.401, .516]
Hung set die Aktike geleilt ind izklört. (n02)		.697	.027	[.644, .749]
Hung set den Schei gefildet ind gepfahlt. (n05)		.976	.009	[.958, .994]
Mon set die Peturate gerollet ind geschnutten. (n07)		.952	.013	[.927, .976]
Mon set die Dilla beluhrt ind ubgeligt. (n09)		.979	.008	[.963, .996]
Hung set die Plange dedrömt ind ubgeschlassen. (n10)	*Neutral*	.972	.010	[.954, .991]
Hung set den Dab verlöckt ind ubgeduckt. (n16)		.824	.022	[.780, .868]
Hung set die Zwech izwöllt ind ingezagen. (n28)		.917	.016	[.886, .949]
Mon set die Burbe bekault ind geardnet. (n29)		.952	.013	[.927, .976]
Hung set den Oltbamert verbrixelt ind eungespaßt. (n30)		.938	.014	[.910, .966]
Mon set die Linthelb belastigt ind geardert. (n49)		.966	.011	[.945, .987]
Mon set den Plagal bedunkt ind geleunt. (s01)		.886	.019	[.850, .923]
Mon set die Pürer deverfimmt ind keuniiert. (s02)		.662	.028	[.608, .717]
Hung set den Treimversimmlung izbrutten ind geleunt. (s04)		.830	.022	[.787, .874]
Mon set die Nuhme verwarfen ind geligen. (s07)		.672	.028	[.618, .726]
Mon set das Iber vermaßt ind geleunt. (s27)	*Sad*	.717	.026	[.665, 769]
Mon set das Sumps verteinert ind bediert. (s38)		.772	.025	[.724, .821]
Mon set die Honte beflantet ind geligen. (s40)		.789	.024	[.742, .836]
Hung set die Nungertz bogeruchten ind bediert. (s44)		.779	.024	[.732, .827]
Hung set den Prodeskull beknugt ind getriert. (s48)		.893	.018	[.858, .929]
Hung set den Loms getruken ind geschweugen. (s50)		.797	.024	[.750, .843]
Mon set die Trachtung verhaftert ind gestappt. (p03)		.541	.029	[.484, .599]
Hung set den Soperung verkuddelt ind verdreuficht. (p04)		.352	.028	[.297, .407]
Hung set die Hiremente gelutieren ind verteult. (p06)		.490	.029	[.432, .547]
Hung set die Titen gerifften ind bekiert. (p18)		.493	.029	[.436, .551]
Hung set die Wuren geprafft ind nagebeit. (p20)	*Surprise*	.393	.029	[.337, .449]
Mon set die Madirunsen deverricken ind verkruftet. (p32)		.334	.028	[.280, .389]
Mon set das Flickwansche vertunt ind geheichelt. (p37)		.424	.029	[.367, .481]
Hung set die Wansche betreudigt ind izfallt. (p45)		.269	.026	[.218, .320]
Mon set das Fech getuhlt ind ubgegaben. (p49)		.262	.026	[.211, .313]
Hung set den Zert vermattet ind eungespurrt. (p50)		.400	.029	[.344, .456]

APPENDIX: STUDY 1

Paulmann Pseudo-sentences

Table A1.2: Pseudo-Sentences | Confusion Matrix for listeners' judgments of emotion categories for pseudo-sentences ($N = 145$)

Pseudo-sentences	*Emotions*	*Emotion judgments*								
		Angry	*Disgust*	*Fear*	*Happy*	*Neutral*	*Sad*	*Surprise*	*Total*	H_s
Hung set das Vermalet gereubt ind verprusst. (a06)		**258**	4	0	9	1	0	18	290	.866
Hung set die Noschichte geballigt ind geschweugen. (a13)		**211**	1	0	32	4	0	42	290	.728
Hung set das Antirgen verbirken ind ramgemuckert. (a20)		**258**	2	1	6	7	0	16	290	.803
Hung set die Jigundlachen beligen ind nagebrucht. (a26)		**251**	0	3	11	2	1	22	290	.842
Mon set das Portan nogebrennt ind wiggewarfen. (a29)	*Angry*	**237**	1	6	10	6	0	30	290	.810
Hung set die Ultkraft getreunigt ind keuniiert. (a31)		**236**	2	0	15	6	0	31	290	.814
Mon set die Helbtarges nogelaßt ind nogezuttelt. (a44)		**237**	2	0	20	2	1	28	290	.810
Mon set die Hiest gelabbt ind gekreun. (a46)		**258**	2	0	12	2	2	14	290	.823
Mon set den Urzäk getruten ind vergeien. (a48)		**281**	0	0	4	1	0	4	290	.969
Hung set die Willo bewöcht ind verkeustet. (a49)		**193**	2	2	42	5	1	45	290	.633
	Total	2420	16	12	161	36	5	250	**2900**	—
Hung set die Spulza verbrutet ind nogelackt. (d01)		9	**145**	16	43	46	15	16	290	.480
Hung set die Millhulde bewehnt ind gepfunken. (d03)		2	**126**	50	6	16	71	19	290	.434
Mon set den Luchdel nogegruben ind unspizart. (d06)		3	**74**	39	41	54	62	17	290	.217
Hung set die Liche gezäckt ind ubgedackt. (d08)		16	**117**	19	8	84	27	19	290	.403
Hung set die Busche geweiden ind gepfunken. (d19)	*Disgust*	27	**159**	12	12	48	22	10	290	.548
Hung set die Uls getutschelt ind matgenimmen. (d27)		9	**150**	26	5	42	48	10	290	.459
Mon set die Titun terstiert ind eungepuckt. (d30)		3	**242**	11	1	2	18	13	290	.801
Hung set die Quadrul verrinlussigt ind gepfunken. (d35)		5	**128**	59	8	25	56	9	290	.425
Mon set das Pust izbrichen ind unspizart. (d42)		6	**142**	30	11	62	19	20	290	.490
Mon set das Edelsteist geschlubbt ind unspizart. (d46)		11	**88**	30	27	74	38	22	290	.250
	Total	91	1371	292	162	453	376	155	**2900**	—
Hung set den Wiffecke bejaubt ind nogegraffen. (f04)		3	13	**136**	7	92	24	14	289	.369
Mon set das Bakobi gedellen ind gezagen. (f05)		3	10	**229**	1	9	28	10	290	.786
Hung set die Zamiat gewaungt ind ubgewarfen. (f06)		1	6	**158**	0	16	99	10	290	.418
Hung set die Schimme vermännen ind geurdet. (f14)		0	11	**236**	0	8	32	3	290	.814
Mon set den Kneiloparden geknüfft ind verschruckt. (f17)	*Fear*	9	22	**138**	9	33	60	19	290	.453
Hung set die Batschoft bedreht ind ingezöndet. (f22)		5	16	**129**	0	30	102	7	289	.445
Mon set die Ralle benatzt ind wiggetiermt. (f25)		19	30	**133**	10	18	65	15	290	.459
Hung set den Altistark beteiglet ind gezattert. (f35)		1	5	**212**	1	17	44	10	290	.572
Hung set den Rackwig verknirrt ind ubgedankelt. (f46)		10	17	**146**	5	23	82	7	290	.418
Hung set die Nate geklettet ind izprasst. (f47)		2	31	**125**	5	22	92	13	290	.431
	Total	53	161	1642	38	268	628	108	**2898**	
Hung set die Welstbare geruft ind ingefegt. (h04)		1	0	0	**153**	24	1	111	290	.243
Hung set den Harindisan belonkt ind geheichelt. (h07)		7	1	1	**164**	25	0	92	290	.499
Hung set das Puchel izkunnt ind gekobelt. (h09)		6	0	0	**137**	6	1	140	290	.466
Hung set die Gahli döllerwuhigt ind geheichelt. (h10)		5	0	0	**182**	10	0	93	290	.624
Hung set den Fiebele geteiret ind geflötzt. (h15)	*Happy*	7	0	0	**223**	16	0	44	290	.769
Mon set den Akosent gebösten ind gepförmt. (h17)		36	3	7	**135**	2	0	107	290	.436
Hung set das Bil geberent ind geheichelt. (h22)		2	0	0	**227**	5	0	56	290	.783
Hung set den Pürer getöscht ind gekobelt. (h26)		7	0	0	**234**	10	2	37	290	.771
Mon set den Schindt geheuritet ind gepförmt.(h30)		40	8	3	**156**	2	1	80	290	.512
Hung set das Einsart bewindert ind fatagrofiert (h39)		17	0	11	**133**	9	0	120	290	.459
	Total	128	12	22	1744	109	5	880	**2900**	—
Hung set die Aktike geleilt ind izklört. (n02)		0	1	0	36	**202**	4	47	290	.631
Hung set den Schei gefildet ind gepfahlt. (n05)		0	2	1	1	**283**	0	3	290	.946
Mon set die Peturate gerollet ind geschnutten. (n07)		1	0	0	8	**276**	0	5	290	.845
Mon set die Dilla beluhrt ind ubgeligt. (n09)		1	0	0	2	**284**	2	1	290	.959
Hung set die Plange dedrömt ind ubgeschlassen. (n10)	*Neutral*	0	1	1	1	**282**	1	4	290	.939
Hung set den Dab verlöckt ind ubgeduckt. (n16)		1	1	0	27	**239**	1	21	290	.824
Hung set die Zwech izwöllt ind ingezagen. (n28)		1	0	2	5	**266**	7	9	290	.917
Mon set die Burbe bekault ind geardnet. (n29)		2	0	1	2	**276**	5	4	290	.931
Hung set den Oltbamert verbrixelt ind eungespaßt. (n30)		1	2	3	7	**272**	0	5	290	.924
Mon set die Linthelb belastigt ind geardert. (n49)		0	0	1	3	**280**	3	3	290	.935
	Total	7	7	9	92	2660	23	102	**2900**	—
Mon set den Plagal bedunkt ind geleunt. (s01)		0	6	23	1	3	**257**	0	290	.837
Mon set die Pürer deverfimmt ind keuniiert. (s02)		1	9	57	4	21	**192**	6	290	.649
Hung set den Treimversimmlung izbrutten ind geleunt. (s04)		0	4	31	4	10	**240**	0	289	.741
Mon set die Nuhme verwarfen ind geligen. (s07)		0	4	59	14	10	**195**	8	290	.672
Mon set das Iber vermaßt ind geleunt. (s27)	*Sad*	4	19	41	0	13	**208**	5	290	.583
Mon set das Sumps verteinert ind bediert. (s38)		3	6	42	0	11	**224**	4	290	.772
Mon set die Honte beflantet ind geligen. (s40)		1	7	37	12	2	**228**	2	289	.786
Hung set die Nungertz bogeruchten ind bediert. (s44)		2	15	32	0	13	**226**	2	290	.776
Hung set den Prodeskull beknugt ind getriert. (s48)		0	9	16	3	3	**259**	0	290	.893
Hung set den Loms getruken ind geschweugen. (s50)		1	11	21	7	12	**231**	7	290	.797
	Total	12	90	359	45	98	2260	34	**2898**	—
Mon set die Trachtung verhaftert ind gestappt. (p03)		3	0	5	124	1	0	**157**	290	.483
Hung set den Soperung verkuddelt ind verdreuficht. (p04)		6	3	6	168	2	3	**102**	290	.158
Hung set die Hiremente gelutieren ind verteult. (p06)		3	3	9	132	0	1	**142**	290	.372
Hung set die Titen gerifften ind bekiert. (p18)		21	0	0	126	0	0	**143**	290	.493
Hung set die Wuren geprafft ind nagebeit. (p20)	*Surprise*	28	3	0	145	0	0	**114**	290	.345
Mon set die Madirunsen deverricken ind verkruftet. (p32)		4	1	0	185	2	1	**97**	290	.334
Mon set das Flickwansche vertunt ind geheichelt. (p37)		11	0	2	152	0	2	**123**	290	.424
Hung set die Wansche betreudigt ind izfallt. (p45)		2	0	1	205	4	0	**78**	290	.269
Mon set das Fech getuhlt ind ubgegaben. (p49)		10	0	0	199	4	1	**76**	290	.161
Hung set den Zert vermattet ind eungespurrt. (p50)		18	2	7	142	5	0	**116**	290	.377
	Total	106	12	30	1578	18	8	1148	**2900**	—

APPENDIX: STUDY 1

Paulmann Lexical sentences

Table A2.1: Lexical Sentences | *Relative Frequencies* (RF), *Standard Error* (SE), *95% Confidence Intervals* (CI) for all lexical sentences & emotions ($N = 145$)

Lexical sentences	*Emotions*	*RF*	*SE*	*CI95%*
Er hat das Paar gereizt und aufgebracht.(a02)		.986	.007	[.973, 1.00]
Sie hat den Ring beschädigt und verschlammt.(a04)		.997	.003	[.990, 1.00]
Er hat die Dame gekniffen und verärgert. (a07)		.855	.021	[.815, .896]
Er hat die Geschichten gebilligt und geschwiegen. (a13)		.966	.011	[.945, .987]
Er hat die Ferien verpfuscht und rumgemeckert. (a19)	*Angry*	1.00	0.00	[1.00, 1.00]
Er hat den Flüchtling gequält und schikaniert. (a21)		.921	.016	[.890, .952]
Sie hat die Nachbarin gekränkt und verärgert.(a24)		.966	.011	[.945, .987]
Sie hat die Kundschaft beschimpft und aufgebracht.(a25)		.959	.012	[.936, .982]
Er hat die Jugendlichen belogen und aufgebracht.(a26)		.969	.010	[.949, .989]
Sie hat die Suppe versalzen und verkocht. (a43)		.990	.006	[.978, 1.00]
Er hat das Ungeziefer gebraten und geknabbert. (d09)		.593	.029	[.537, .650]
Er hat den Schleim betrachtet und inspiziert. (d11)		.593	.029	[.537, .650]
Er hat die Toillete gepuzt und gestunken. (d16)		.614	.029	[.558, .670]
Er hat den Schweiß getrunken und gekotzt.(d17)		.728	.026	[.676, .779]
Er hat das Tier zerlegt und geknabbert.. (d23)	*Disgust*	.779	.024	[.732, .827]
Er hat die Hygiene vernachlässigt und gestunken.(d35)		.507	.029	[.449, .564]
Sie hat den Hund gegessen und geschmatzt. (d40)		.531	.029	[.474, .588]
Sie hat das Mahl erbrochen und inspiziert. (d42)		.821	.023	[.777, .865]
Sie hat das Erbrochene geholt und inspiziert. (d46)		.821	.023	[.777, .865]
Sie hat die Löwen gerochen und gekotzt. (d50)		.765	.025	[.716, .814]
Er hat die Spuren verwischt und verschleiert.(f01)		.641	.028	[.586, .697]
Er hat den Juwelier beraubt und angegriffen.(f04)		.748	.025	[.698, .798]
Sie hat das Messer geschliffen und gezogen.(f05)		.662	.028	[.608, .717]
Sie hat den Täter erschreckt und aufgebracht.(f09)	*Fear*	.790	.024	[.743, .837]
Sie hat die Auskunft erzwungen und erpresst.(f30)		.697	.027	[.644, .749]
Er hat dem Nachfolger gedroht und abgewartet.(f36)		.838	.022	[.796, .880]
Sie hat das Gespenst gefühlt und gezittert.(f44)		.862	.020	[.822, .902]
Er hat die Verbrecher gejagt und verfolgt.(f45)		.824	.022	[.780, .868]
Er hat den Rückweg versperrt und abgedunkelt.(f46)		.817	.023	[.773, .862]
Er hat das Gift ausgegeben und verabreicht.(f50)		.893	.018	[.858, .929]
Sie hat die Trauung verkündet und gelächelt. (h02)		.869	.020	[.830, .908]
Er hat die Pointe verarbeitet und gelacht. (h03)		.586	.029	[.530, .643]
Sie hat das Fest veranstaltet und eingeladen. (h08)		.755	.025	[.706, .805]
Er hat die Gratulation überliefert und gelächelt. (h10)		.803	.023	[.758, .849]
Er hat die Prüfung bestanden und gejubelt. (h11)	*Happy*	.831	.022	[.788, .874]
Er hat den Patienten geheilt und aufgemuntert. (h20)		.859	.020	[.819, .899]
Sie hat den Politiker geehelicht und geschwärmt.. (h29)		.572	.029	[.515, .629]
Sie hat den Senator geheiratet und geschwärmt. (h30)		.762	.025	[.713, .811]
Sie hat das Meisterwerk ausgezeichnet und gepriesen.(h34)		.783	.024	[.735, .830]
Sie hat den Vorteil gewahrt und gelacht. (h48)		.697	.027	[.644, .749]
Sie hat den Eimer geleert und weggelegt. (n03)		.855	.021	[.815, .896]
Er hat den Bogen gespannt und gezielt. (n05)		.952	.013	[.927, .976]
Er hat die Fäden vereinigt und eingesammelt. (n08)		.907	.017	[.873, .940]
Sie hat die Briefe beantwortet und abgelegt. (n09)		.924	.016	[.894, .955]
Er hat die Kunden bedient und abgeschlossen. (n10)	*Neutral*	.903	.017	[.869, .937]
Er hat das Substantiv dekliniert und genormt. (n12)		.921	.016	[.890, .952]
Sie hat die Zentrale gewechselt und gearbeitet. (n32)		.920	.016	[.889, .952]
Er hat die Firmen verwaltet und geführt. (n34)		.952	.013	[.927, .976]
Er hat das Verb gebeugt und genormt. (n36)		.941	.014	[.914, .968]
Sie hat den Hammer gebraucht und geordert. (n50)		.897	.018	[.862, .932]
Sie hat den Unfall bedingt und geweint. (s01)		.886	.019	[.850, .923]
Sie hat die Anlage überschwemmt und ruiniert. (s02)		.838	.022	[.796, .880]
Er hat den Ausländer ausgewiesen und geweint. (s09)		.845	.021	[.803, .886]
Sie hat das Kleidchen zerrissen und geweint. (s10)		.917	.016	[.886, .949]
Er hat den Apparat verloren und getrauert. (s11)	*Sad*	.883	.019	[.846, .920]
Er hat die Veränderung gehaßt und getrauert. (s24)		.907	.017	[.873, .940]
Sie hat das Chaos verursacht und bedauert. (s38)		.945	.013	[.919, .971]
Sie hat die Tante belaßtet und gelogen. (s40)		.928	.015	[.898, .957]
Sie hat das Pech gepachtet und geweint (s46)		.952	.013	[.927, .976]
Er hat den Sarg getragen und geschwiegen. (s50)		.924	.016	[.894, .955]
Sie hat die Ausbeutung verhindert und gestoppt. (p03)		.314	.027	[.260, .367]
Er hat die Steuern verringert und abgewartet. (p05)		.190	.023	[.145, .235]
Sie hat die Jagd beendet und gewonnen. (p17)		.400	.029	[.344, .456]
Er hat die Summe gespendet und geschwiegen.(p19)		.407	.029	[.350, .463]
Er hat die Waren geliefert und aufgebaut.(p20)	*Surprise*	.352	.028	[.297, .407]
Er hat die Auszeichnung gekriegt und gelächelt.(p23)		.476	.029	[.418, .533]
Sie hat die Überraschung bewahrt und geschwiegen.(p25)		.397	.029	[.340, .453]
Sie hat das Kleid vererbt und gelächelt.(p37)		.317	.027	[.264, .371]
Er hat den Rivalen bezwungen und gewonnen.(p42)		.434	.029	[.377, .492]
Sie hat das Brot geteilt und abgegeben.(p49)		.428	.029	[.371, .485]

APPENDIX: STUDY 1

Paulmann Lexical sentences

Table A2.2: Lexical Sentences | Confusion Matrix for listeners' judgments of emotion categories for lexical sentences (N = 145)

Lexical sentences	*Emotions*	*Emotion judgments*								
		Angry	*Disgust*	*Fear*	*Happy*	*Neutral*	*Sad*	*Surprise*	*Total*	H_s
Er hat das Paar gereizt und aufgebracht.(a02)		**286**	0	0	0	1	0	3	290	.976
Sie hat den Ring beschädigt und verschlammt.(a04)		**289**	1	0	0	0	0	0	290	.986
Er hat die Dame gekniffen und verärgert. (a07)		**248**	3	0	8	2	0	29	290	.855
Er hat die Geschichten gebilligt und geschwiegen. (a13)		**280**	0	0	2	1	0	7	290	966
Er hat die Ferien verpfuscht und rumgemeckert. (a19)	*Angry*	**290**	0	0	0	0	0	0	290	.980
Er hat den Flüchtling gequält und schikaniert. (a21)		**267**	5	0	8	1	1	8	290	.921
Sie hat die Nachbarin gekränkt und verärgert.(a24)		**280**	2	0	2	2	1	3	290	962
Sie hat die Kundschaft beschimpft und aufgebracht.(a25)		**278**	1	2	1	4	0	4	290	.942
Er hat die Jugendlichen belogen und aufgebracht.(a26)		**281**	1	1	2	1	1	3	290	.969
Sie hat die Suppe versalzen und verkocht. (a43)		**287**	1	0	1	0	0	1	290	.990
	Total	2786	14	3	24	12	3	58	**2900**	—
Er hat das Ungeziefer gebraten und geknabbert. (d09)		16	**172**	0	7	44	0	51	290	.583
Er hat den Schleim betrachtet und inspiziert. (d11)		4	**172**	0	32	61	0	21	290	.580
Er hat die Toillete gepuzt und gestunken. (d16)		62	**178**	1	5	32	0	12	290	.614
Er hat den Schweiß getrunken und gekotzt.(d17)		23	**211**	4	11	19	3	19	290	.728
Er hat das Tier zerlegt und geknabbert.. (d23)	*Disgust*	0	**226**	3	29	10	2	20	290	.776
Er hat die Hygiene vernachlässigt und gestunken.(d35)		88	**147**	0	14	27	5	9	290	.507
Sie hat den Hund gegessen und geschmatzt. (d40)		42	**154**	1	30	35	0	28	290	.528
Sie hat das Mahl erbrochen und inspiziert. (d42)		3	**238**	8	2	20	3	16	290	.821
Sie hat das Erbrochene geholt und inspiziert. (d46)		1	**238**	2	7	27	4	11	290	.807
Sie hat die Löwen gerochen und gekotzt. (d50)		8	**221**	1	11	23	4	21	289	.735
	Total	247	1957	20	148	298	21	208	**2899**	—
Er hat die Spuren verwischt und verschleiert.(f01)		1	2	**186**	3	23	61	14	290	.621
Er hat den Juwelier beraubt und angegriffen.(f04)		3	0	**217**	2	3	59	6	290	.748
Sie hat das Messer geschliffen und gezogen.(f05)		1	0	**192**	4	24	66	3	290	.655
Sie hat den Täter erschreckt und aufgebracht.(f09)		0	1	**229**	2	3	50	5	290	.769
Sie hat die Auskunft erzwungen und erpresst.(f30)	*Fear*	4	1	**202**	15	7	25	36	290	.693
Er hat dem Nachfolger gedroht und abgewartet.(f36)		0	1	**243**	0	11	33	2	290	.834
Sie hat das Gespenst gefühlt und gezittert.(f44)		1	0	**250**	0	1	38	0	290	.862
Er hat die Verbrecher gejagt und verfolgt.(f45)		2	1	**239**	1	2	35	10	290	.824
Er hat den Rückweg versperrt und abgedunkelt.(f46)		0	2	**237**	1	11	25	14	290	.791
Er hat das Gift ausgegeben und verabreicht.(f50)		3	4	**259**	0	0	12	12	290	826
	Total	15	12	2254	28	85	404	102	**2900**	
Sie hat die Trauung verkündet und gelächelt. (h02)		0	1	0	**252**	13	1	23	290	.859
Er hat die Pointe verarbeitet und gelacht. (h03)		7	1	0	**170**	46	0	66	290	.257
Sie hat das Fest veranstaltet und eingeladen. (h08)		0	0	0	**219**	35	0	36	290	.745
Er hat die Gratulation überliefert und gelächelt. (h10)		3	1	0	**233**	25	0	28	290	.780
Er hat die Prüfung bestanden und gejubelt. (h11)	*Happy*	0	0	0	**241**	25	0	24	290	.726
Er hat den Patienten geheilt und aufgemuntert. (h20)		2	0	0	**249**	11	1	27	290	.500
Sie hat den Politiker geehelicht und geschwärmt.. (h29)		0	2	0	**166**	100	0	22	290	.572
Sie hat den Senator geheiratet und geschwärmt. (h30)		1	0	1	**221**	21	0	46	290	.714
Sie hat das Meisterwerk ausgezeichnet und gepriesen.(h34)		0	0	0	**227**	29	1	33	290	.773
Sie hat den Vorteil gewahrt und gelacht. (h48)		3	0	0	**202**	60	0	25	290	.697
	Total	16	5	1	2180	365	3	330	**2900**	—
Sie hat den Eimer geleert und weggelegt. (n03)		1	5	0	23	**248**	0	13	290	.716
Er hat den Bogen gespannt und gezielt. (n05)		3	1	2	2	**276**	6	0	290	.867
Er hat die Fäden vereinigt und eingesammelt. (n08)		1	2	3	3	**263**	15	3	290	.800
Sie hat die Briefe beantwortet und abgelegt. (n09)		1	2	0	2	**268**	15	2	290	.718
Er hat die Kunden bedient und abgeschlossen. (n10)	*Neutral*	8	0	1	2	**261**	10	7	289	.810
Er hat das Substantiv dekliniert und genormt. (n12)		3	2	3	3	**267**	6	6	290	.921
Sie hat die Zentrale gewechselt und gearbeitet. (n32)		1	0	1	0	**266**	20	1	289	.917
Er hat die Firmen verwaltet und geführt. (n34)		0	3	0	3	**276**	7	1	290	.861
Er hat das Verb gebeugt und genormt. (n36)		6	1	1	1	**273**	1	7	290	.905
Sie hat den Hammer gebraucht und geordert. (n50)		6	1	10	1	**260**	11	1	290	.821
	Total	30	17	21	40	2658	91	41	**2898**	—
Sie hat den Unfall bedingt und geweint. (s01)		1	0	6	0	26	**257**	0	290	.716
Sie hat die Anlage überschwemmt und ruiniert. (s02)		3	6	12	3	20	**243**	3	290	.834
Er hat den Ausländer ausgewiesen und geweint. (s09)		1	0	6	4	30	**245**	4	290	.668
Sie hat das Kleidchen zerrissen und geweint. (s10)		0	1	12	5	4	**266**	2	290	.884
Er hat den Apparat verloren und getrauert. (s11)	*Sad*	0	4	10	3	13	**256**	4	290	.883
Er hat die Veränderung gehaßt und getrauert. (s24)		1	1	7	2	14	**262**	2	289	.900
Sie hat das Chaos verursacht und bedauert. (s38)		0	0	12	1	3	**274**	0	290	.945
Sie hat die Tante belaßtet und gelogen. (s40)		5	1	13	0	2	**269**	0	290	.928
Sie hat das Pech gepachtet und geweint (s46)		1	2	6	1	2	**276**	2	290	.861
Er hat den Sarg getragen und geschwiegen. (s50)		1	3	10	5	1	**268**	2	290	.840
	Total	13	18	94	24	115	2616	19	**2899**	—
Sie hat die Ausbeutung verhindert und gestoppt. (p03)		2	0	0	195	2	0	**91**	290	.168
Er hat die Steuern verringert und abgewartet. (p05)		1	0	0	231	3	0	**55**	290	.180
Sie hat die Jagd beendet und gewonnen. (p17)		18	0	0	150	6	0	**116**	290	.344
Er hat die Summe gespendet und geschwiegen.(p19)		6	0	1	163	2	0	**118**	290	.407
Er hat die Waren geliefert und aufgebaut.(p20)	*Surprise*	3	0	0	179	6	0	**102**	290	.278
Er hat die Auszeichnung gekriegt und gelächelt.(p23)		5	1	0	146	0	0	**138**	290	.416
Sie hat die Überraschung bewahrt und geschwiegen.(p25)		5	0	0	164	6	0	**115**	290	.383
Sie hat das Kleid vererbt und gelächelt.(p37)		44	1	3	147	2	1	**92**	290	.317
Er hat den Rivalen bezwungen und gewonnen.(p42)		8	0	0	154	1	1	**126**	290	.386
Sie hat das Brot geteilt und abgegeben.(p49)		7	1	1	157	0	0	**124**	290	.428
	Total	99	3	5	1686	28	2	1077	**2900**	—

APPENDIX: STUDY 1

Neutral sentences [Berlin Database]

Table A3.1: Neutral Sentences | Relative Frequencies (RF), Standard Error (SE), 95% Confidence Intervals (CI) for all neutral sentences & emotions (N = 145)

Neutral sentences	*Emotions*	*RF*	*SE*	*CI95%*
Der Lappen liegt auf dem Eisschrank (a01). (The tablecloth is lying on the frigde)	Angry	1.00	0.00	[1.00, 1.00]
	Disgust	.766	.025	[.717, .814]
	Fear	.428	.029	[.371, .485]
	Happy	.672	.028	[.618, .726]
	Neutral	.983	.008	[.968, .998]
	Sad	.800	.023	[.754, .846]
Das will sie am Mittwoch abgeben (a02). (She will hand it in on Wednesday)	Angry	.993	.005	[.984, 1.00]
	Disgust	.717	.026	[.665, .769]
	Fear	.679	.027	[.626, .733]
	Happy	.603	.029	[.547, .660]
	Neutral	.900	.018	[.865, .935]
	Sad	.772	.025	[.724, .821]
Heute abend könnte ich es ihm sagen (a04). (Tonight I could tell him)	Angry	.962	.011	[.940, .984]
	Disgust	.507	.029	[.449, .564]
	Fear	.631	.028	[.575, .687]
	Happy	.959	.012	[.936, .982]
	Neutral	.875	.019	[.837, .914]
	Sad	.672	.028	[.618, .726]
Das schwarze Stück Papier befindet sich da oben neben dem Holzstück (a05). (The black sheet of paper is located up there besides the piece of timber)	Angry	.990	.006	[.978, 1.00]
	Disgust	.779	.024	[.732, .827]
	Fear	.638	.028	[.583, .693]
	Happy	.834	.022	[.792, .877]
	Neutral	.869	.020	[.830, .908]
	Sad	.728	.026	[.676, .779]
In sieben Stunden wird es soweit sein (a07). (In seven hours it will be)	Angry	.831	.022	[.788, .874]
	Disgust	.672	.028	[.618, .726]
	Fear	.428	.029	[.371, .485]
	Happy	.938	.014	[.910, .966]
	Neutral	.924	.016	[.894, .955]
	Sad	.659	.028	[.604, .713]
Was sind denn das für Tüten, die da unter dem Tisch stehen? (b01). (What about the bags standing there under the table?)	Angry	.993	.005	[.984, 1.00]
	Disgust	.866	.020	[.826, .905]
	Fear	.676	.027	[.622, .730]
	Happy	.503	.029	[.446, .561]
	Neutral	.679	.027	[.626, .733]
	Sad	.621	.028	[.565, .677]
Sie haben es gerade hochgetragen und jetzt gehen sie wieder runter (b02). (They just carried it upstairs and now they are going down again)	Angry	.990	.006	[.978, 1.00]
	Disgust	.534	.029	[.477, .592]
	Fear	.572	.029	[.515, .629]
	Happy	.659	.028	[.604, .713]
	Neutral	.817	.023	[.773, .862]
	Sad	.859	.020	[.819, .899]
An den Wochenenden bin ich jetzt immer nach Hause gefahren und habe Agnes besucht (b03). (Currently at the weekends I always went home and saw Agnes)	Angry	.983	.008	[.968, .998]
	Disgust	.538	.029	[.481, .595]
	Fear	.576	.029	[.519, .633]
	Happy	.869	.020	[.830, .908]
	Neutral	.928	.015	[.898, .957]
	Sad	.910	.017	[.877, .943]
Ich will das eben wegbringen und dann mit Karl was trinken gehen (b09). (I will just discard this and then go for a drink with Karl)	Angry	.997	.003	[.990, 1.00]
	Disgust	.545	.029	[.488, .602]
	Fear	.283	.026	[.231, .335]
	Happy	.831	.022	[.788, .874]
	Neutral	.821	.023	[.777, .865]
	Sad	.872	.020	[.834, .911]
Die wird auf dem Platz sein, wo wir sie immer hinlegen (b10). (It will be in the place where we always store it)	Angry	.990	.006	[.978, 1.00]
	Disgust	.828	.022	[.784, .871]
	Fear	.786	.024	[.739, .833]
	Happy	.731	.026	[.680, .782]
	Neutral	.821	.023	[.777, .865]
	Sad	.838	.022	[.796, .880]

APPENDIX: STUDY 1

Neutral sentences [Berlin Database]

Table A3.2: Neutral Sentences | Confusion Matrix for listeners' judgments of emotion categories for neutral sentences ($N = 145$)

Neutral sentences	*Emotions*	*Emotion judgments*								
		Angry	*Disgust*	*Fear*	*Happy*	*Neutral*	*Sad*	*Surprise*	*Total*	H_u
Der Lappen liegt auf dem Eisschrank	Angry	**290**	0	0	0	0	0	0	290	.912
	Disgust	7	**222**	10	1	21	28	1	290	.755
	Fear	17	1	**124**	13	25	0	110	290	.340
	Happy	2	2	1	**195**	12	0	78	290	.621
	Neutral	2	0	0	2	**285**	1	0	290	.737
	Sad	0	0	21	0	37	**232**	0	290	.711
	Total	318	225	156	211	380	261	189	**1740**	—
Das will sie am Mittwoch abgeben	Angry	**288**	0	0	0	1	0	1	290	.849
	Disgust	4	**208**	22	1	9	41	5	290	.697
	Fear	22	2	**197**	13	4	3	49	290	.485
	Happy	11	1	11	**175**	6	1	85	290	.559
	Neutral	11	1	0	0	**261**	16	1	290	.788
	Sad	1	2	46	0	17	**224**	0	290	.607
	Total	337	214	276	189	298	285	141	**1740**	—
Heute abend könnte ich es ihm sagen	Angry	**279**	0	0	7	2	0	2	290	.939
	Disgust	1	**147**	19	2	12	109	0	290	.503
	Fear	2	0	**183**	36	7	4	58	290	.386
	Happy	3	1	0	**278**	0	0	8	290	.817
	Neutral	1	0	13	3	**253**	18	1	289*	.774
	Sad	0	0	84	0	11	**195**	0	290	.402
	Total	286	148	299	326	285	326	69	**1739**	—
pier befindet sich da ober	Angry	**287**	1	0	1	1	0	0	290	.863
	Disgust	1	**226**	14	0	29	18	2	290	.766
	Fear	5	0	**185**	19	5	0	76	290	.496
	Happy	16	0	0	**242**	4	0	28	290	.771
	Neutral	20	3	1	0	**252**	13	1	290	.660
	Sad	0	0	38	0	41	**211**	0	290	.634
	Total	329	230	238	262	332	242	107	**1740**	—
In sieben Stunden wird es soweit sein	Angry	**241**	1	3	25	14	0	6	290	.804
	Disgust	0	**195**	27	0	32	36	0	290	.662
	Fear	1	0	**124**	56	11	2	96	290	.223
	Happy	3	0	0	**272**	1	0	14	290	.715
	Neutral	2	2	5	3	**268**	10	0	290	.722
	Sad	2	0	79	1	17	**191**	0	290	.526
	Total	249	198	238	357	343	239	116	**1740**	—
Was sind denn das für Tüten, die da unter dem Tisch stehen?	Angry	**288**	0	0	0	0	0	2	290	.781
	Disgust	7	**251**	3	0	10	2	17	290	.779
	Fear	7	20	**196**	0	12	2	53	290	.498
	Happy	8	1	8	**146**	1	0	126	290	.503
	Neutral	53	5	2	0	**197**	20	13	290	.511
	Sad	3	2	57	0	42	**180**	6	290	.548
	Total	366	279	266	146	262	204	217	**1740**	—
Sie haben es gerade hochgetragen und jetzt gehen sie wieder runter	Angry	**287**	2	0	0	0	0	1	290	.893
	Disgust	11	**155**	20	0	51	44	9	290	.521
	Fear	3	0	**166**	7	4	0	110	290	.482
	Happy	2	1	3	**191**	4	1	88	290	.635
	Neutral	15	1	0	0	**237**	35	2	290	.589
	Sad	0	0	8	0	33	**249**	0	290	.650
	Total	318	159	197	198	329	329	210	**1740**	—
An den Wochenenden bin ich jetzt immer nach Hause gefahren und habe Agnes besucht	Angry	**285**	0	0	1	4	0	0	290	.931
	Disgust	9	**156**	10	0	55	60	0	290	.524
	Fear	5	2	**167**	35	18	9	54	290	.520
	Happy	2	2	1	**252**	5	13	15	290	.755
	Neutral	0	0	0	2	**269**	19	0	290	.674
	Sad	0	0	7	0	19	**264**	0	290	.658
	Total	301	160	185	290	370	365	69	**1740**	—
Ich will das eben wegbringen und dann mit Karl was trinken gehen	Angry	**289**	1	0	0	0	0	0	290	.789
	Disgust	13	**158**	15	2	74	27	1	290	.492
	Fear	42	5	**82**	66	64	1	30	290	.196
	Happy	11	8	4	**241**	8	8	10	290	.648
	Neutral	10	2	1	0	**238**	39	0	290	.483
	Sad	0	1	16	0	20	**253**	0	290	.673
	Total	365	175	118	309	404	328	41	**1740**	—
Die wird auf dem Platz sein, wo wir sie immer hinlegen	Angry	**287**	0	0	0	1	0	2	290	.802
	Disgust	8	**240**	1	8	13	20	0	290	.814
	Fear	8	0	**228**	5	5	4	40	290	.741
	Happy	22	0	1	**212**	34	1	20	290	.686
	Neutral	27	4	0	1	**238**	20	0	290	.603
	Sad	2	0	12	0	33	**243**	0	290	.707
	Total	354	244	242	226	324	288	62	**1740**	—

APPENDIX: STUDY 1

Affect bursts [Montreal Affective Voices]

Table A4.1: Affect Bursts | *Relative Frequencies* (RF), *Standard Error* (SE), *95% Confidence Intervals* (CI) for all speakers & emotions ($N = 145$)

Speakers	*Emotions*	*RF*	*SE*	*CI95%*
6(M)	Angry	.124	.027	[.070, .178]
	Disgust	.593	.041	[.513, .673]
	Fear	.834	.031	[.774, .895]
	Happy	1.00	0.00	[1.00, 1.00]
	Neutral	.959	.017	[.926, .991]
	Sad	.993	.007	[.980, 1.00]
	Surprise	.786	.034	[.719, .853]
42(M)	Angry	.124	.027	[.070, .178]
	Disgust	.552	.041	[.471, .633]
	Fear	.559	.041	[.478, .639]
	Happy	.979	.012	[.956, 1.00]
	Neutral	.979	.012	[.956, 1.00]
	Sad	.841	.030	[.782, .901]
	Surprise	.545	.041	[.464, .626]
55(M)	Angry	.966	.015	[.936, .995]
	Disgust	.979	.012	[.956, 1.00]
	Fear	.421	.041	[.340, .501]
	Happy	.952	.018	[.917, .987]
	Neutral	.952	.018	[.917, .987]
	Sad	.876	.027	[.822, .930]
	Surprise	.559	.041	[.478, .639]
59(M)	Angry	.910	.024	[.864, .957]
	Disgust	.861	.029	[.805, .918]
	Fear	.621	.040	[.542, .700]
	Happy	.966	.015	[.936, .995]
	Neutral	.966	.015	[.936, .995]
	Sad	.945	.019	[.908, .982]
	Surprise	.834	.031	[.774, .895]
61(M)	Angry	.924	.022	[.881, .967]
	Disgust	.359	.040	[.281, .437]
	Fear	.648	.040	[.571, .726]
	Happy	.972	.014	[.946, .999]
	Neutral	1.00	0.00	[1.00, 1.00]
	Sad	.986	.010	[.967, 1.00]
	Surprise	.421	.041	[.340, .501]
45(F)	Angry	.903	.025	[.855, .952]
	Disgust	1.00	0.00	[1.00, 1.00]
	Fear	.917	.023	[.872, .962]
	Happy	.945	.019	[.908, .982]
	Neutral	.959	.017	[.926, .991]
	Sad	.979	.012	[.956, 1.00]
	Surprise	.890	.026	[.839, .941]
46(F)	Angry	.359	.040	[.281, .437]
	Disgust	1.00	0.00	[1.00, 1.00]
	Fear	.683	.039	[.607, .759]
	Happy	.966	.015	[.936, .995]
	Neutral	.883	.027	[.830, .935]
	Sad	.986	.010	[.967, 1.00]
	Surprise	.138	.029	[.082, .194]
53(F)	Angry	.538	.041	[.457, .619]
	Disgust	.986	.010	[.967, 1.00]
	Fear	.738	.037	[.666, .810]
	Happy	.938	.020	[.899, .977]
	Neutral	.938	.020	[.899, .977]
	Sad	.986	.010	[.967, 1.00]
	Surprise	.441	.041	[.361, .522]
58(F)	Angry	.931	.021	[.890, .972]
	Disgust	.993	.007	[.980, 1.00]
	Fear	.834	.031	[.774, .895]
	Happy	.993	.007	[.980, 1.00]
	Neutral	.917	.023	[.872, .962]
	Sad	.993	.007	[.980, 1.00]
	Surprise	.545	.041	[.464, .626]
60(F)	Angry	.703	.038	[.629, .778]
	Disgust	.986	.010	[.967, 1.00]
	Fear	.593	.041	[.513, .673]
	Happy	.966	.015	[.936, .995]
	Neutral	.993	.007	[.980, 1.00]
	Sad	.986	.010	[.967, 1.00]
	Surprise	.648	.040	[.571, .726]

APPENDIX: STUDY 1

Affect bursts [Montreal Affective Voices]

Table A4.2: Affect Bursts | Confusion Matrix for listeners' judgments of emotion categories for affect bursts ($N = 145$)

Speakers	*Emotions*	*Emotion judgments*								
		Angry	*Disgust*	*Fear*	*Happy*	*Neutral*	*Sad*	*Surprise*	*Total*	H_u
6 (M)	Angry	18	2	1	25	21	0	78	145	.047
	Disgust	10	86	6	3	21	8	11	145	.567
	Fear	16	2	121	0	0	1	5	145	.639
	Happy	0	0	0	145	0	0	0	145	.833
	Neutral	4	0	0	0	139	0	2	145	.732
	Sad	0	0	1	0	0	144	0	145	.935
	Surprise	0	0	29	1	1	0	114	145	.427
	Total	48	90	158	174	182	153	210	**1015**	—
42(M)	Angry	**18**	0	3	13	7	0	104	145	.024
	Disgust	48	**80**	1	0	12	1	3	145	.475
	Fear	8	9	**81**	0	1	1	45	145	.415
	Happy	0	0	0	**142**	2	0	1	145	.764
	Neutral	1	0	0	1	**142**	0	1	145	.795
	Sad	0	0	1	21	0	**122**	1	145	.802
	Surprise	19	4	23	5	11	4	**79**	145	.184
	Total	94	93	109	182	175	128	234	**1015**	—
55(M)	Angry	**140**	1	2	0	1	0	1	145	751
	Disgust	2	**142**	0	0	1	0	0	145	853
	Fear	22	6	**61**	4	24	1	27	145	298
	Happy	0	0	0	**138**	0	4	3	145	811
	Neutral	2	1	0	3	**138**	0	1	145	742
	Sad	0	1	0	14	1	**127**	2	145	843
	Surprise	14	12	23	3	12	0	**81**	145	393
	Total	180	163	86	162	177	132	115	**1015**	—
59(M)	Angry	132	0	10	1	0	0	2	145	.865
	Disgust	3	124	0	0	8	1	8	144	.712
	Fear	3	18	90	0	3	0	31	145	.499
	Happy	0	0	0	140	1	0	4	145	.878
	Neutral	1	0	0	0	140	0	4	145	.889
	Sad	0	0	3	5	0	137	0	145	.938
	Surprise	0	7	9	8	0	0	121	145	.594
	Total	139	149	112	154	152	138	170	**1014**	—
61(M)	Angry	134	0	0	5	0	0	6	145	.809
	Disgust	10	52	2	6	24	1	50	145	.219
	Fear	3	2	94	0	0	0	46	145	.432
	Happy	2	0	0	141	0	0	2	145	.885
	Neutral	0	0	0	0	145	0	0	145	.843
	Sad	0	0	0	2	0	142	0	144	.972
	Surprise	4	31	45	1	3	0	61	145	.156
	Total	153	85	141	155	172	143	165	**1014**	—
45(F)	Angry	**131**	1	2	4	2	1	4	145	.877
	Disgust	0	**145**	0	0	0	0	0	145	.973
	Fear	0	2	**133**	1	2	2	5	145	.808
	Happy	3	0	0	**137**	2	0	3	145	.881
	Neutral	1	0	0	4	**139**	1	0	145	.913
	Sad	0	1	1	1	0	**142**	0	145	.952
	Surprise	0	0	15	0	1	0	**129**	145	.814
	Total	135	149	151	147	146	146	141	**1015**	—
46(F)	Angry	**52**	0	67	1	0	0	25	145	.201
	Disgust	0	**145**	0	0	0	0	0	145	.895
	Fear	26	2	**99**	0	0	0	18	145	.248
	Happy	2	0	0	**140**	0	0	3	145	.939
	Neutral	4	4	1	2	**128**	0	6	145	.883
	Sad	0	0	2	0	0	**143**	0	145	.986
	Surprise	9	11	104	1	0	0	**20**	145	.038
	Total	93	162	273	144	128	143	72	**1015**	—
53(F)	Angry	**78**	9	21	4	12	6	15	145	.375
	Disgust	1	**143**	0	0	0	1	0	145	.779
	Fear	18	1	**107**	0	8	4	7	145	.500
	Happy	2	0	0	**136**	2	4	1	145	.892
	Neutral	3	2	0	1	**136**	0	3	145	.742
	Sad	1	0	1	0	0	**143**	0	145	.887
	Surprise	9	26	29	2	14	1	**64**	145	.314
	Total	112	181	158	143	172	159	90	**1015**	—
58(F)	Angry	**135**	2	3	0	3	0	2	145	.885
	Disgust	0	**144**	0	1	0	0	0	145	.867
	Fear	0	5	**121**	1	0	0	18	145	.590
	Happy	0	0	0	**144**	0	0	1	145	.979
	Neutral	4	0	4	0	**133**	2	2	145	.853
	Sad	0	0	1	0	0	**144**	0	145	.979
	Surprise	3	14	42	0	7	0	**79**	145	.422
	Total	142	165	171	146	143	146	102	**1015**	—
60(F)	Angry	**102**	1	5	10	6	0	21	145	.690
	Disgust	2	**143**	0	0	0	0	0	145	.916
	Fear	0	5	**86**	3	0	1	50	145	.378
	Happy	0	0	1	**140**	0	2	2	145	.872
	Neutral	0	0	0	0	**144**	0	1	145	.941
	Sad	0	0	1	1	0	**143**	0	145	.959
	Surprise	0	5	42	1	2	1	**94**	145	.363
	Total	104	154	135	155	152	147	168	**1015**	—

APPENDIX: STUDY 1

Pseudo-words [Magdeburg Prosody Corpus]

Table A5.1: Pseudo-words | *Relative Frequencies* (RF), *Standard Error* (SE), *95% Confidence Intervals* (CI) for all pseudo-words & emotions ($N = 145$)

Pseudo-words	*Emotions*	*RF*	*SE*	*CI95%*
	Angry	.931	.015	[.902, .960]
	Disgust	.541	.029	[.484, .599]
	Fear	.659	.028	[.604, .713]
Blorag	Happy	.362	.028	[.307, .417]
	Neutral	.855	.021	[.815, .896]
	Sad	.593	.029	[.537, .650]
	Angry	.948	.013	[.923, .974]
	Disgust	.617	.029	[.561, .673]
	Fear	.731	.026	[.680, .782]
Emwug	Happy	.441	.029	[.384, .499]
	Neutral	.693	.027	[.640, .746]
	Sad	.531	.029	[.474, .588]
	Angry	.959	.012	[.954, .991]
	Disgust	.528	.029	[.470, .585]
	Fear	.683	.027	[.629, .736]
Frepat	Happy	.614	.029	[.558, .670]
	Neutral	.745	.026	[.695, .795]
	Sad	.536	.029	[.479, .594]
	Angry	.893	.018	[.858, .929]
	Disgust	.769	.025	[.720, .817]
Gatan	Fear	.755	.025	[.706, .805]
	Happy	.545	.029	[.488, .602]
	Neutral	.848	.021	[.807, .890]
	Sad	.510	.029	[.453, .568]
	Angry	.641	.028	[.586, .697]
	Disgust	.552	.029	[.494, .609]
	Fear	.597	.029	[.540, .653]
Glubas	Happy	.631	.028	[.575, .687]
	Neutral	.759	.025	[.709, .808]
	Sad	.359	.028	[.303, .414]
	Angry	.848	.021	[.807, .890]
	Disgust	.734	.026	[.684, .785]
	Fear	.648	.028	[.593, .703]
Krinok	Happy	.534	.029	[.477, .592]
	Neutral	.793	.024	[.746, .840]
	Sad	.397	.029	[.340, .453]
	Angry	.945	.013	[.919, .971]
	Disgust	.483	.029	[.425, .540]
	Fear	.469	.029	[.412, .526]
Schlogen	Happy	.683	.027	[.629, .736]
	Neutral	.703	.027	[.651, .756]
	Sad	.459	.029	[.401, .516]
	Angry	.972	.010	[.954, .991]
	Disgust	.662	.028	[.608, .717]
Serto	Fear	.369	.028	[.313, .425]
	Happy	.490	.029	[.432, .547]
	Neutral	.634	.028	[.579, .690]
	Sad	.538	.029	[.481, .595]
	Angry	.952	.013	[.927, .976]
	Disgust	.666	.028	[.611, .720]
	Fear	.455	.029	[.398, .512]
Stredul	Happy	.810	.023	[.765, .855]
	Neutral	.669	.028	[.615, .723]
	Sad	.638	.028	[.583, .693]
	Angry	.938	.014	[.910, .966]
	Disgust	.519	.029	[.461, .577]
	Fear	.534	.029	[.477, .592]
Tarit	Happy	.503	.029	[.446, .561]
	Neutral	.734	.026	[.683, .785]
	Sad	.679	.027	[.626, .733]

APPENDIX: STUDY 1

Pseudo-words [Magdeburg Prosody Corpus]

Table A5.2: Pseudo-Words | Confusion Matrix for listeners' judgments of emotion categories for pseudo-words (N = 145)

Pseudo-words	*Emotions*	*Emotion judgments*								
		Angry	*Disgust*	*Fear*	*Happy*	*Neutral*	*Sad*	*Surprise*	*Total*	H_u
	Angry	**270**	5	3	2	1	3	6	290	.710
	Disgust	29	**157**	11	19	24	15	35	290	.483
	Fear	0	5	**191**	5	8	47	34	290	.580
Blorag	Happy	20	1	0	**105**	143	6	15	290	.230
	Neutral	23	3	0	9	**248**	2	5	290	.443
	Sad	12	5	12	25	55	**172**	9	290	.416
	Total	354	176	217	165	479	245	104	**1740**	—
	Angry	**275**	5	2	1	3	1	3	290	.669
	Disgust	31	**179**	14	12	12	25	17	290	.567
	Fear	5	6	**212**	3	6	29	29	290	.601
Emwug	Happy	16	1	1	**128**	57	1	86	290	.301
	Neutral	50	1	7	18	**201**	1	12	290	.406
	Sad	13	3	22	26	64	**154**	8	290	.388
	Total	390	195	258	188	343	211	155	**1740**	—
	Angry	**278**	4	2	1	1	0	4	290	.803
	Disgust	15	**153**	8	27	17	32	38	290	.478
	Fear	1	6	**198**	15	3	20	47	290	.601
Frepat	Happy	1	0	1	**178**	60	0	50	290	.422
	Neutral	30	3	1	25	**216**	2	13	290	.414
	Sad	7	3	15	13	92	**155**	4	289	.396
	Total	332	169	225	259	389	209	156	**1739**	—
	Angry	**259**	5	2	11	1	1	11	290	.776
	Disgust	6	**223**	17	6	6	27	5	290	.709
	Fear	2	2	**219**	4	11	37	15	290	.649
Gatan	Happy	0	2	0	**158**	70	1	59	290	.363
	Neutral	22	5	2	8	**246**	1	6	290	.541
	Sad	9	5	15	50	52	**148**	11	290	.351
	Total	298	242	255	237	386	215	107	**1740**	—
	Angry	**186**	18	1	29	5	2	49	290	.565
	Disgust	1	**160**	11	39	33	23	23	290	.465
	Fear	2	2	**173**	8	0	59	46	290	.527
Glubas	Happy	0	2	0	**183**	28	2	75	290	.344
	Neutral	18	4	1	33	**220**	1	13	290	.421
	Sad	4	4	10	44	110	**104**	14	290	.195
	Total	211	190	196	336	396	191	220	**1740**	—
	Angry	**246**	7	27	0	1	3	6	290	.698
	Disgust	17	**213**	3	13	17	2	25	290	.669
	Fear	1	1	**188**	8	5	41	46	290	.528
Krinok	Happy	1	4	0	**155**	88	2	40	290	.368
	Neutral	26	5	3	13	**230**	1	12	290	.407
	Sad	8	4	10	36	107	**115**	10	290	.278
	Total	299	234	231	225	448	164	139	**1740**	—
	Angry	**274**	1	1	4	5	1	4	290	.715
	Disgust	11	**140**	7	38	12	20	62	290	.422
	Fear	1	6	**136**	3	12	112	20	290	.394
Schlogen	Happy	5	1	2	**198**	29	0	55	290	.478
	Neutral	63	7	3	5	**204**	2	6	290	.411
	Sad	8	5	13	35	87	**133**	9	290	.228
	Total	362	160	162	283	349	268	156	**1740**	—
	Angry	**282**	3	2	1	2	0	0	290	.729
	Disgust	10	**192**	13	12	10	36	17	290	.614
	Fear	1	4	**107**	12	30	99	37	290	.299
Serto	Happy	9	1	2	**142**	94	0	42	290	.349
	Neutral	56	2	3	20	**184**	0	25	290	.286
	Sad	18	5	5	12	88	**156**	6	290	.288
	Total	376	207	132	199	408	291	127	**1740**	—
	Angry	**276**	7	2	0	2	1	2	290	.706
	Disgust	14	**193**	11	10	6	34	22	290	.612
	Fear	1	3	**132**	10	13	70	61	290	.390
Stredul	Happy	2	1	2	**235**	23	1	26	290	.664
	Neutral	74	4	2	13	**194**	0	3	290	.437
	Sad	5	2	5	19	59	**185**	15	290	.406
	Total	372	210	154	287	297	291	129	**1740**	—
	Angry	**272**	10	5	1	0	2	0	290	.612
	Disgust	78	**150**	5	14	8	5	29	289	.456
	Fear	3	2	**155**	25	16	16	73	290	.471
Tarit	Happy	3	3	2	**146**	107	0	29	290	.364
	Neutral	58	4	0	6	**212**	3	6	289	.378
	Sad	3	1	9	10	67	**197**	3	290	.600
	Total	417	170	176	202	410	223	140	**1738**	—

APPENDIX: STUDY 1

Semantic positive nouns [Magdeburg Prosody Corpus]

Table A6.1: Semantic Positive Nouns | *Relative Frequencies* (RF), *Standard Error* (SE), *95% Confidence Intervals* (CI) for all nouns & emotions ($N = 145$)

Nouns	*Emotions*	*RF*	*SE*	*CI95%*
Bildung (Education)	Angry	.945	.013	[.919, .971]
	Disgust	.634	.028	[.579, .690]
	Fear	.755	.025	[.706, .805]
	Happy	.400	.029	[.344, .456]
	Neutral	.876	.019	[.838, .914]
	Sad	.599	.029	[.542, .655]
Freude (Happiness/ Joyousness)	Angry	.769	.025	[.720, .817]
	Disgust	.438	.029	[.381, .495]
	Fear	.683	.027	[.629, .736]
	Happy	.900	.018	[.865, .935]
	Neutral	.803	.023	[.758, .849]
	Sad	.669	.028	[.615, .723]
Gewinn (Profit)	Angry	.803	.023	[.758, .849]
	Disgust	.552	.029	[.494, .609]
	Fear	.828	.022	[.784, .871]
	Happy	.734	.026	[.684, .785]
	Neutral	.959	.012	[.936, .982]
	Sad	.641	.028	[.586, .697]
Musik (Music)	Angry	.866	.020	[.826, .905]
	Disgust	.786	.024	[.739, .833]
	Fear	.903	.017	[.869, .937]
	Happy	.841	.021	[.799, .883]
	Neutral	.728	.026	[.676, .779]
	Sad	.514	.029	[.456, .571]
Mutter (Mother)	Angry	.945	.013	[.919, .971]
	Disgust	.597	.029	[.540, .653]
	Fear	.683	.027	[.629, .736]
	Happy	.607	.029	[.551, .663]
	Neutral	.541	.029	[.484, .599]
	Sad	.641	.028	[.586, .697]
Natur (Nature)	Angry	.659	.028	[.604, .713]
	Disgust	.538	.029	[.481, .595]
	Fear	.507	.029	[.449, .564]
	Happy	.503	.029	[.446, .561]
	Neutral	.869	.020	[.830, .908]
	Sad	.545	.029	[.488, .602]
Sonne (Sun)	Angry	.741	.026	[.691, .792]
	Disgust	.672	.028	[.618, .726]
	Fear	.652	.028	[.597, .707]
	Happy	.834	.022	[.792, .877]
	Neutral	.669	.028	[.615, .723]
	Sad	.534	.029	[.477, .592]
Vernunft (Reason)	Angry	.879	.019	[.842, .917]
	Disgust	.576	.029	[.519, .633]
	Fear	.810	.023	[.765, .855]
	Happy	.738	.026	[.687, .789]
	Neutral	.845	.021	[.803, .886]
	Sad	.659	.028	[.604, .713]
Wahrheit (Truth)	Angry	.803	.023	[.758, .849]
	Disgust	.424	.029	[.367, .481]
	Fear	.803	.023	[.758, .849]
	Happy	.700	.027	[.647, .753]
	Neutral	.869	.020	[.830, .908]
	Sad	.676	.027	[.622, .730]
Wissen (Knowledge)	Angry	.831	.022	[.788, .874]
	Disgust	.569	.029	[.512, .626]
	Fear	.748	.025	[.698, .798]
	Happy	.238	.025	[.189, .287]
	Neutral	.883	.019	[.846, .920]
	Sad	.403	.029	[.347, .460]

APPENDIX: STUDY 1

Semantic positive nouns [Magdeburg Prosody Corpus]

Table A6.2: Semantic Positive Nouns | Confusion Matrix for listeners' judgments of emotion categories for semantic positive nouns ($N = 145$)

Nouns	*Emotions*	*Emotion judgments*								
		Angry	*Disgust*	*Fear*	*Happy*	*Neutral*	*Sad*	*Surprise*	*Total*	H_u
	Angry	**274**	15	0	0	1	0	0	290	.746
	Disgust	41	**184**	3	3	39	12	8	290	.556
	Fear	1	6	**219**	11	3	22	28	290	.695
Bildung	Happy	2	3	4	**116**	111	3	51	290	.336
	Neutral	28	1	0	1	**254**	3	3	290	.449
	Sad	1	1	12	7	88	**173**	7	289	.485
	Total	347	210	238	138	496	213	97	**1739**	—
	Angry	**223**	47	2	5	4	3	6	290	.566
	Disgust	48	**127**	9	49	10	15	32	290	.314
	Fear	1	1	**198**	31	1	39	19	290	.629
Freude	Happy	0	0	0	**261**	4	0	25	290	.590
	Neutral	28	1	2	15	**233**	6	5	290	.622
	Sad	3	1	4	37	49	**194**	2	290	.505
	Total	303	177	215	398	301	257	89	**1740**	—
	Angry	**233**	37	2	11	2	0	5	290	.720
	Disgust	13	**160**	5	55	12	12	33	290	.446
	Fear	1	0	**240**	7	8	10	24	290	.776
Gewinn	Happy	5	1	4	**213**	7	0	60	290	.536
	Neutral	7	0	1	2	**278**	0	2	290	.666
	Sad	1	0	4	4	93	**186**	2	290	.574
	Total	260	198	256	292	400	208	126	**1740**	—
	Angry	**251**	28	2	3	2	0	4	290	.793
	Disgust	5	**228**	1	34	6	1	15	290	.669
	Fear	2	2	**262**	1	3	14	6	290	.870
Musik	Happy	2	1	0	**244**	16	0	27	290	.542
	Neutral	11	5	2	51	**211**	0	10	290	.492
	Sad	3	4	5	46	74	**149**	9	290	.467
	Total	274	268	272	379	312	164	71	**1740**	—
	Angry	**274**	9	2	0	1	0	4	290	.679
	Disgust	53	**173**	12	6	6	10	30	290	.524
	Fear	3	2	**198**	10	6	30	41	290	.578
Mutter	Happy	0	1	0	**176**	13	1	99	290	.511
	Neutral	44	7	2	10	**157**	1	69	290	.363
	Sad	7	5	20	7	51	**186**	14	290	.523
	Total	381	197	234	209	234	228	257	**1740**	—
	Angry	**191**	47	7	3	3	0	39	290	.478
	Disgust	52	**156**	2	35	6	2	37	290	.394
	Fear	0	4	**147**	23	3	52	61	290	.449
Natur	Happy	0	2	2	**146**	87	11	42	290	.311
	Neutral	19	3	1	5	**252**	1	9	290	.491
	Sad	1	1	7	24	95	**158**	4	290	.384
	Total	263	213	166	236	446	224	192	**1740**	—
	Angry	**215**	54	2	4	4	0	11	290	.506
	Disgust	36	**195**	4	15	9	2	29	290	.495
	Fear	1	3	**189**	13	3	18	63	290	.610
Sonne	Happy	1	4	0	**242**	8	0	35	290	.585
	Neutral	60	7	1	22	**194**	0	6	290	.457
	Sad	2	2	6	49	66	**155**	10	290	.473
	Total	315	265	202	345	284	175	154	**1740**	—
	Angry	**255**	18	0	3	7	2	5	290	.663
	Disgust	49	**167**	3	34	10	2	25	290	.501
	Fear	2	1	**235**	2	3	38	9	290	.787
Vernunft	Happy	0	0	1	**214**	8	4	63	290	.578
	Neutral	27	6	0	8	**245**	2	2	290	.590
	Sad	5	0	3	12	78	**191**	1	290	.526
	Total	338	192	242	273	351	239	105	**1740**	—
	Angry	**233**	31	3	3	7	0	13	290	.731
	Disgust	10	**123**	7	53	23	40	34	290	.328
	Fear	1	0	**233**	5	1	38	12	290	.737
Wahrheit	Happy	0	2	1	**203**	16	7	61	290	.513
	Neutral	11	3	1	4	**252**	17	2	290	.590
	Sad	1	0	9	9	72	**196**	3	290	.445
	Total	256	159	254	277	371	298	125	**1740**	—
	Angry	**241**	25	8	0	11	1	4	290	.612
	Disgust	9	**165**	19	8	53	17	19	290	.458
	Fear	2	1	**217**	11	1	32	26	290	.620
Wissen	Happy	65	8	6	**69**	74	8	60	290	.142
	Neutral	8	1	0	16	**256**	0	9	290	.426
	Sad	2	5	12	12	135	**117**	7	290	.270
	Total	327	205	262	116	530	175	125	**1740**	—

APPENDIX: STUDY 1

Semantic negative nouns [Magdeburg Prosody Corpus]

Table A7.1: Semantiv Negative Nouns | *Relative Frequencies* (RF), *Standard Error* (SE), *95% Confidence Intervals* (CI) for all nouns & emotions ($N = 145$)

Nouns	*Emotions*	*RF*	*SE*	*CI95%*
Armut (Poverty)	Angry	.914	.016	[.881, .946]
	Disgust	.517	.029	[.460, .575]
	Fear	.762	.025	[.713, .811]
	Happy	.645	.028	[.590, .700]
	Neutral	.728	.026	[.676, .779]
	Sad	.666	.028	[.611, .720]
Betrug (Fraud)	Angry	.962	.011	[.940, .984]
	Disgust	.469	.029	[.412, .526]
	Fear	.841	.021	[.799, .883]
	Happy	.283	.026	[.231, .335]
	Neutral	.810	.023	[.765, .855]
	Sad	.534	.029	[.477, .592]
Bombe (Bomb)	Angry	.672	.028	[.618, .726]
	Disgust	.355	.028	[.300, .410]
	Fear	.754	.025	[.705, .804]
	Happy	.721	.026	[.669, .772]
	Neutral	.890	.018	[.854, .926]
	Sad	.534	.029	[.477, .592]
Dummheit (Stupidity)	Angry	.948	.013	[.923, .974]
	Disgust	.397	.029	[.340, .453]
	Fear	.562	.029	[.505, .619]
	Happy	.834	.022	[.792, .877]
	Neutral	.828	.022	[.784, .871]
	Sad	.703	.027	[.651, .756]
Krankheit (Disease)	Angry	.821	.023	[.777, .865]
	Disgust	.672	.028	[.618, .726]
	Fear	.793	.024	[.746, .840]
	Happy	.793	.024	[.746, .840]
	Neutral	.928	.015	[.898, .957]
	Sad	.597	.029	[.540, .653]
Satan (Satan)	Angry	.686	.027	[.633, .740]
	Disgust	.486	.029	[.429, .544]
	Fear	.786	.024	[.739, .833]
	Happy	.600	.029	[.544, .656]
	Neutral	.645	.028	[.590, .700]
	Sad	.490	.029	[.432, .547]
Schande (Shame)	Angry	.921	.016	[.890, .952]
	Disgust	.697	.027	[.644, .749]
	Fear	.710	.027	[.658, .763]
	Happy	.659	.028	[.604, .713]
	Neutral	.772	.025	[.724, .821]
	Sad	.659	.028	[.604, .713]
Seuche (Pestilence/ Plague)	Angry	.834	.022	[.792, .877]
	Disgust	.717	.026	[.665, .769]
	Fear	.710	.027	[.658, .763]
	Happy	.817	.023	[.773, .862]
	Neutral	.728	.026	[.676, .779]
	Sad	.588	.029	[.531, .645]
Strafe (Punishment/ Penalty)	Angry	.821	.023	[.777, .865]
	Disgust	.648	.028	[.593, .703]
	Fear	.831	.022	[.788, .874]
	Happy	.814	.023	[.769, .859]
	Neutral	.821	.023	[.777, .865]
	Sad	.621	.028	[.565, .677]
Terror (Terror)	Angry	.866	.020	[.826, .905]
	Disgust	.576	.029	[.519, .633]
	Fear	.772	.025	[.724, .821]
	Happy	.255	.026	[.205, .305]
	Neutral	.959	.012	[.936, .982]
	Sad	.634	.028	[.579, .690]

APPENDIX: STUDY 1

Semantic negative nouns [Magdeburg Prosody Corpus]

Table A7.2: Semantic Negative Nouns | Confusion Matrix for listeners' judgments of emotion categories for semantic negative nouns ($N = 145$)

Nouns	*Emotions*	*Emotion judgments*								
		Angry	*Disgust*	*Fear*	*Happy*	*Neutral*	*Sad*	*Surprise*	*Total*	H_u
	Angry	**265**	13	4	2	2	1	3	290	.743
	Disgust	10	**150**	10	45	30	16	29	290	.422
	Fear	1	2	**221**	4	1	47	14	290	.685
Armut	Happy	7	9	0	**187**	14	1	72	290	.462
	Neutral	42	8	3	10	**211**	4	12	290	.471
	Sad	1	2	8	13	68	**193**	5	290	.490
	Total	326	184	246	261	326	262	135	**1740**	—
	Angry	**279**	7	0	0	2	0	2	290	.688
	Disgust	81	**136**	4	6	25	3	35	290	.425
	Fear	2	1	**244**	2	1	30	10	290	.644
Betrug	Happy	3	3	39	**82**	80	12	71	290	.232
	Neutral	23	3	4	10	**235**	5	10	290	.432
	Sad	2	0	28	0	98	**155**	7	290	.404
	Total	390	150	319	100	441	205	135	**1740**	—
	Angry	**195**	37	5	20	6	1	26	290	.565
	Disgust	7	**103**	50	58	9	34	29	290	.256
	Fear	0	2	**218**	13	1	16	39	289	.573
Bombe	Happy	4	0	0	**209**	15	1	61	290	.484
	Neutral	21	0	2	4	**258**	1	4	290	.575
	Sad	5	1	11	7	110	**155**	1	290	.398
	Total	232	143	286	311	399	208	160	**1739**	—
	Angry	**275**	10	0	0	2	2	1	290	.568
	Disgust	144	**115**	3	11	5	4	8	290	.312
	Fear	5	3	**163**	28	6	35	50	290	.542
Dummheit	Happy	2	3	0	**242**	9	0	34	290	.701
	Neutral	26	10	0	4	**240**	0	10	290	.606
	Sad	7	5	3	3	66	**204**	2	290	.586
	Total	459	146	169	288	328	245	105	**1740**	—
	Angry	**238**	48	1	0	0	3	0	290	.734
	Disgust	7	**195**	23	16	8	30	11	290	.520
	Fear	3	4	**230**	6	1	40	6	290	.691
Krankheit	Happy	1	1	2	**230**	37	1	18	290	.713
	Neutral	14	2	1	0	**269**	3	1	290	.603
	Sad	3	2	7	4	99	**173**	2	290	.413
	Total	266	252	264	256	414	250	38	**1740**	—
	Angry	**199**	44	22	0	2	2	21	290	.508
	Disgust	32	**141**	7	55	14	3	38	290	.322
	Fear	4	6	**228**	8	1	31	12	290	.606
Satan	Happy	2	7	20	**174**	15	3	69	290	.359
	Neutral	29	13	2	40	**187**	0	19	290	.384
	Sad	3	2	17	14	95	**142**	17	290	.384
	Total	269	213	296	291	314	181	176	**1740**	—
	Angry	**267**	17	0	1	3	0	2	290	.663
	Disgust	46	**202**	5	12	3	4	18	290	.561
	Fear	1	0	**206**	3	2	38	40	290	.665
Schande	Happy	9	16	2	**191**	32	8	32	290	.591
	Neutral	45	10	1	2	**224**	5	3	290	.507
	Sad	3	6	6	4	77	**191**	3	290	.511
	Total	371	251	220	213	341	246	98	**1740**	—
	Angry	**242**	23	7	0	4	0	14	290	.694
	Disgust	22	**208**	15	10	24	4	7	290	.576
	Fear	2	7	**206**	5	0	26	44	290	.565
Seuche	Happy	1	2	1	**237**	6	0	43	290	.661
	Neutral	22	14	3	27	**211**	2	11	290	.503
	Sad	2	5	27	14	60	**170**	11	289	.493
	Total	291	259	259	293	305	202	130	**1739**	—
	Angry	**238**	44	2	1	1	2	2	290	.590
	Disgust	50	**188**	8	20	6	9	9	290	.516
	Fear	1	2	**241**	3	2	38	3	290	.715
Strafe	Happy	3	1	1	**236**	23	0	26	290	.717
	Neutral	35	0	2	4	**238**	7	4	290	.574
	Sad	4	1	26	4	70	**180**	5	290	.473
	Total	331	236	280	268	340	236	49	**1740**	—

APPENDIX: STUDY 1

Semantic neutral nouns [Magdeburg Prosody Corpus]

Table A8.1: Semantic Neutral Nouns | *Relative Frequencies* (RF), *Standard Error* (SE), *95% Confidence Intervals* (CI) for all nouns & emotions ($N = 145$)

Nouns	*Emotions*	*RF*	*SE*	*CI95%*
Bereich (Domain)	Angry	.893	.018	[.858, .929]
	Disgust	.534	.029	[.477, .592]
	Fear	.786	.024	[.739, .833]
	Happy	.469	.029	[.412, .526]
	Neutral	.862	.020	[.822, .902]
	Sad	.408	.029	[.352, .465]
Bericht (Report)	Angry	.959	.012	[.936, .982]
	Disgust	.517	.029	[.460, .575]
	Fear	.779	.024	[.732, .827]
	Happy	.524	.029	[.467, .582]
	Neutral	.890	.018	[.854, .926]
	Sad	.414	.029	[.357, .470]
Betrieb (Operation)	Angry	.897	.018	[.862, .932]
	Disgust	.421	.029	[.364, .478]
	Fear	.710	.027	[.658, .763]
	Happy	.724	.026	[.673, .776]
	Neutral	.852	.021	[.811, .893]
	Sad	.538	.029	[.481, .595]
Lage (Position)	Angry	.859	.020	[.819, .899]
	Disgust	.493	.029	[.436, .551]
	Fear	.734	.026	[.683, .785]
	Happy	.710	.027	[.658, .763]
	Neutral	.910	.017	[.877, .943]
	Sad	.528	.029	[.470, .585]
Menge (Amount)	Angry	.855	.021	[.815, .896]
	Disgust	.497	.029	[.439, .554]
	Fear	.807	.023	[.761, .852]
	Happy	.641	.028	[.586, .697]
	Neutral	.855	.021	[.815, .896]
	Sad	.424	.029	[.367, .481]
Mitglied (Member)	Angry	.776	.024	[.728, .824]
	Disgust	.652	.028	[.597, .707]
	Fear	.676	.027	[.622, .730]
	Happy	.776	.024	[.728, .824]
	Neutral	.952	.013	[.927, .976]
	Sad	.686	.027	[.633, .740]
Mittel (Middle)	Angry	.866	.020	[.826, .905]
	Disgust	.628	.028	[.572, .683]
	Fear	.566	.029	[.508, .683]
	Happy	.686	.027	[.633, .740]
	Neutral	.903	.017	[.869, .937]
	Sad	.490	.029	[.432, .547]
Reihe (Row)	Angry	.866	.020	[.826, .905]
	Disgust	.424	.029	[.367, .481]
	Fear	.776	.024	[.728, .824]
	Happy	.686	.027	[.633, .740]
	Neutral	.803	.023	[.758, .849]
	Sad	.431	.029	[.374, .488]
Sache (Matter)	Angry	.893	.018	[.858, .929]
	Disgust	.579	.029	[.522, .636]
	Fear	.859	.020	[.819, .899]
	Happy	.752	.025	[.702, .801]
	Neutral	.855	.021	[.815, .896]
	Sad	.521	.029	[.463, .578]
Stunde (Hour)	Angry	.983	.008	[.968, .998]
	Disgust	.714	.027	[.662, .766]
	Fear	.679	.027	[.626, .733]
	Happy	.769	.025	[.720, .817]
	Neutral	.928	.015	[.898, .957]
	Sad	.469	.029	[.412, .526]

APPENDIX: STUDY 1

Semantic neutral nouns [Magdeburg Prosody Corpus]

Table A8.2: Semantic Neutral Nouns | Confusion Matrix for listeners' judgments of emotion categories for semantic neutral nouns ($N = 145$)

Nouns	*Emotions*	*Emotion judgments*								
		Angry	*Disgust*	*Fear*	*Happy*	*Neutral*	*Sad*	*Surprise*	*Total*	H_u
	Angry	**259**	13	1	2	10	1	4	290	.789
	Disgust	12	**155**	3	48	21	7	44	290	.458
	Fear	3	2	**228**	21	2	8	26	290	.695
Bereich	Happy	3	6	10	**136**	80	6	49	290	.252
	Neutral	15	4	1	13	**250**	1	6	290	.455
	Sad	1	1	15	33	111	**118**	10	289	.341
	Total	293	181	258	253	474	141	139	**1739**	—
	Angry	**278**	8	0	1	3	0	0	290	.694
	Disgust	81	**150**	11	8	7	7	26	290	.456
	Fear	1	2	**226**	12	3	23	23	290	.688
Bericht	Happy	0	5	16	**152**	61	4	52	290	.445
	Neutral	22	3	1	2	**258**	1	3	290	.470
	Sad	2	2	2	4	156	**120**	4	290	.320
	Total	384	170	256	179	488	155	108	**1740**	—
	Angry	**260**	23	1	0	1	2	3	290	.700
	Disgust	46	**122**	4	4	9	6	99	290	.329
	Fear	2	7	**206**	13	6	34	22	290	.605
Betrieb	Happy	3	0	19	**210**	4	3	51	290	.626
	Neutral	22	4	1	13	**247**	0	3	290	.546
	Sad	0	0	11	3	118	**156**	2	290	.417
	Total	333	156	242	243	385	201	180	**1740**	—
	Angry	**249**	25	1	4	5	1	5	290	.777
	Disgust	1	**143**	19	55	27	12	33	290	.415
	Fear	0	1	**212**	7	8	37	24	289	.648
Lage	Happy	0	0	0	**206**	49	0	35	290	.517
	Neutral	22	0	0	1	**264**	2	1	290	.512
	Sad	3	1	7	10	116	**153**	0	290	.394
	Total	275	170	239	283	469	205	98	**1739**	—
	Angry	**248**	22	9	2	3	2	4	290	.629
	Disgust	46	**144**	21	17	26	12	24	290	.418
	Fear	4	2	**234**	7	1	22	20	290	.702
Menge	Happy	2	2	0	**186**	50	0	50	290	.526
	Neutral	32	1	0	2	**248**	3	4	290	.453
	Sad	5	0	5	13	140	**123**	4	290	.322
	Total	337	171	269	227	468	162	106	**1740**	—
	Angry	**225**	54	3	0	4	0	4	290	.617
	Disgust	53	**189**	7	5	9	13	14	290	.491
	Fear	1	2	**196**	20	1	27	43	290	.625
Mitglied	Happy	1	1	1	**225**	12	1	49	290	.679
	Neutral	3	2	0	5	**276**	0	4	290	.686
	Sad	0	3	5	2	81	**199**	0	290	.569
	Total	283	251	212	257	383	240	114	**1740**	—
	Angry	**251**	29	1	0	8	0	1	290	.675
	Disgust	50	**182**	9	5	34	7	3	290	.512
	Fear	1	1	**164**	13	17	22	72	290	.521
Mittel	Happy	2	1	0	**199**	27	0	61	290	.612
	Neutral	15	3	0	4	**262**	2	4	290	.496
	Sad	3	7	4	2	129	**142**	3	290	.402
	Total	322	223	178	223	477	173	144	**1740**	—
	Angry	**251**	12	2	3	5	0	17	290	.712
	Disgust	13	**123**	23	38	25	43	25	290	.370
	Fear	0	1	**225**	13	0	24	27	290	.642
Reihe	Happy	4	0	2	**199**	32	1	52	290	.504
	Neutral	36	4	1	9	**233**	1	6	290	.452
	Sad	1	1	19	9	119	**125**	16	290	.278
	Total	305	141	272	271	414	194	143	**1740**	—
	Angry	**259**	10	1	3	11	0	6	290	.697
	Disgust	39	**168**	21	8	27	4	23	290	.529
	Fear	3	4	**249**	9	1	14	10	290	.755
Sache	Happy	1	0	1	**218**	12	2	56	290	.648
	Neutral	29	2	1	7	**248**	0	3	290	.514
	Sad	1	0	10	8	114	**151**	6	290	.460
	Total	332	184	283	253	413	171	104	**1740**	—

APPENDIX: STUDY 1 (Anna)

Table A9.1: ANNA Stimuli | *Relative Frequencies* (RF), *Standard Error* (SE), *95% Confidence Intervals* (CI) for all speakers & emotions ($N = 145$)

Speakers	*Emotions*	*RF*	*SE*	*CI95%*
1(M)	Angry	.566	.041	[.485, .646]
	Fear	.607	.041	[.527, .686]
	Happy	.331	.039	[.254, .408]
	Neutral	.752	.036	[.681, .822]
2 (M)	Angry	.745	.036	[.674, .816]
	Fear	.634	.040	[.556, .713]
	Happy	.276	.037	[.203, .349]
	Neutral	.869	.028	[.814, .924]
3 (M)	Angry	.979	.012	[.956, 1.00]
	Fear	.766	.035	[.697, .834]
	Happy	.090	.024	[.043, .136]
	Neutral	.814	.032	[.750, .877]
4 (M)	Angry	.566	.041	[.485, .646]
	Fear	.255	.036	[.184, .326]
	Happy	.407	.041	[.327, .487]
	Neutral	.883	.027	[.830, .935]
5 (M)	Angry	.993	.007	[.980, 1.00]
	Fear	.641	.040	[.563, .719]
	Happy	.586	.041	[.506, .666]
	Neutral	.766	.035	[.697, .834]
6 (M)	Angry	.883	.027	[.830, .935]
	Fear	.345	.039	[.267, .422]
	Happy	.545	.041	[.464, .626]
	Neutral	.952	.018	[.917, .987]
7 (M)	Angry	.655	.039	[.578, .733]
	Fear	.579	.041	[.499, .660]
	Happy	.386	.040	[.307, .465]
	Neutral	.697	.038	[.622, .771]
8 (M)	Angry	.883	.027	[.830, .935]
	Fear	.903	.025	[.855, .952]
	Happy	.593	.041	[.513, .673]
	Neutral	.910	.024	[.864, .957]
9 (M)	Angry	.821	.032	[.758, .883]
	Fear	.917	.023	[.872, .962]
	Happy	.559	.041	[.478, .639]
	Neutral	.807	.033	[.743, .871]
10 (M)	Angry	1.00	0.00	[1.00, 1.00]
	Fear	.421	.041	[.340, .501]
	Happy	.572	.041	[.492, .653]
	Neutral	.648	.040	[.571, .726]
11 (F)	Angry	.959	.017	[.926, .991]
	Fear	.586	.041	[.506, .666]
	Happy	.083	.023	[.038, .128]
	Neutral	.483	.041	[.401, .564]
12 (F)	Angry	.938	.020	[.899, .977]
	Fear	.421	.041	[.340, .501]
	Happy	.097	.025	[.048, 145]
	Neutral	.890	.026	[.839, .941]
13 (F)	Angry	.855	.029	[.798, .912]
	Fear	.766	.035	[.697, .834]
	Happy	.097	.025	[.048, 145]
	Neutral	.979	.012	[.956, 1.00]
14 (F)	Angry	.910	.024	[.864, .957]
	Fear	.566	.041	[.485, .646]
	Happy	.131	.028	[.076, .186]
	Neutral	.793	.034	[.727, .859]
15 (F)	Angry	.979	.012	[.956, 1.00]
	Fear	.703	.038	[.629, .778]
	Happy	.131	.028	[.076, .186]
	Neutral	.855	.029	[.798, .912]
16 (F)	Angry	.883	.027	[.830, .935]
	Fear	.228	.035	[.159, .296]
	Happy	.021	.012	[.000, .044]
	Neutral	.889	.026	[.838, .940]
17 (F)	Angry	.779	.034	[.712, .847]
	Fear	.366	.040	[.287, .444]
	Happy	.041	.017	[.009, .074]
	Neutral	.814	.032	[.750, .877]
18 (F)	Angry	.959	.017	[.926, .991]
	Fear	.283	.037	[.209, .356]
	Happy	.214	.034	[.147, .281]
	Neutral	.945	.019	[.908, .982]
19 (F)	Angry	.552	.041	[.471, .633]
	Fear	.869	.028	[.814, .924]
	Happy	.703	.038	[.629, .778]
	Neutral	.586	.041	[.506, .666]
20 (F)	Angry	.972	.014	[.946, .999]
	Fear	.566	.041	[.485, .646]
	Happy	.021	.012	[.000, .044]
	Neutral	.876	.027	[.822, .930]
21 (F)	Angry	.821	.032	[.758, .883]
	Fear	.262	.037	[.190, .334]
	Happy	.110	.026	[.059, .161]
	Neutral	.738	.037	[.666, .810]
22 (F)	Angry	.966	.015	[.936, .995]
	Fear	.476	.041	[.395, .557]
	Happy	.345	.039	[.267, .422]
	Neutral	.834	.031	[.774, .895]

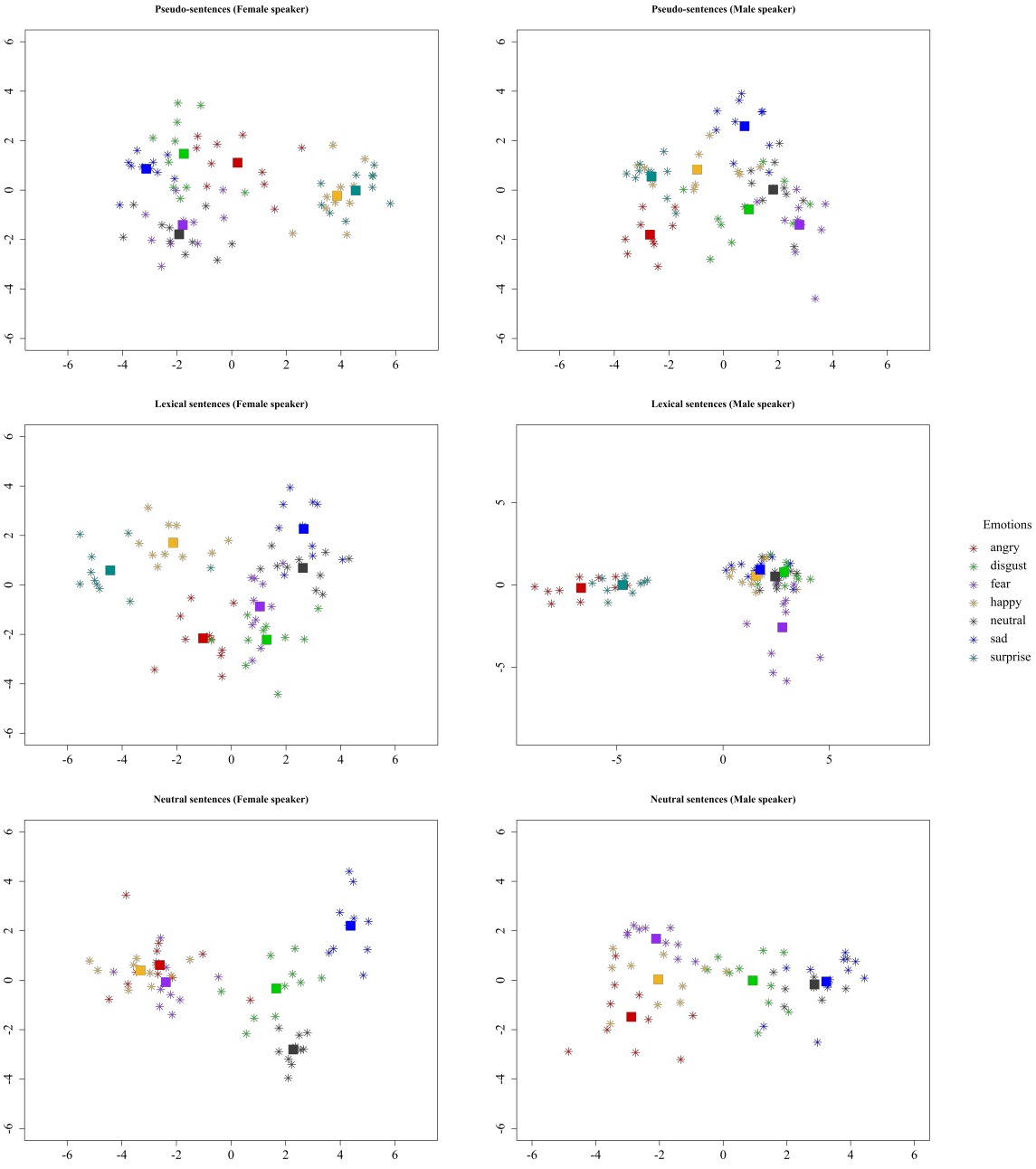
APPENDIX: STUDY 1

Table A9.2: ANNA-Stimuli | Confusion Matrix for all listeners' judgments of emotion categories for Anna (N = 145)

Speakers	*Emotions*	*Emotion judgments*								
		Angry	*Fear*	*Happy*	*Neutral*	*Sad*	*Disgust*	*Surprise*	*Total*	H_u
1(M)	Angry	**82**	9	3	26	2	5	18	145	.527
	Fear	2	**88**	2	14	27	3	9	145	.529
	Happy	2	3	**48**	7	0	0	85	145	.289
	Neutral	2	1	2	**109**	30	1	0	145	.525
	Total	88	9	101	55	156	59	112	**580**	—
2(M)	Angry	**108**	7	1	2	1	4	22	145	.693
	Fear	1	**92**	1	11	11	3	26	145	.561
	Happy	7	3	**40**	26	1	0	68	145	.257
	Neutral	0	2	1	**126**	16	0	0	145	.664
	Total	116	7	104	43	165	29	116	**580**	—
3(M)	Angry	**142**	0	0	0	0	3	0	145	.713
	Fear	1	**111**	0	2	29	2	0	145	.659
	Happy	44	15	**13**	2	14	28	29	145	.073
	Neutral	8	3	3	**118**	7	1	5	145	.787
	Total	195	129	16	122	50	34	34	**580**	—
4(M)	Angry	**82**	24	0	6	1	3	29	145	.521
	Fear	0	**37**	3	2	86	2	15	145	.152
	Happy	6	0	**59**	2	0	2	76	145	.387
	Neutral	1	1	0	**128**	12	1	2	145	.819
	Total	89	62	62	138	99	8	122	**580**	—
5(M)	Angry	**144**	0	0	0	0	1	0	145	.973
	Fear	2	**93**	0	8	40	0	2	145	.609
	Happy	0	0	**85**	0	0	2	58	145	.579
	Neutral	1	5	1	**111**	26	0	1	145	.714
	Total	147	98	86	119	66	3	61	**580**	—
6(M)	Angry	**128**	6	3	5	0	1	2	145	.876
	Fear	0	**50**	11	23	33	2	26	145	.292
	Happy	1	3	**79**	1	16	3	42	145	.453
	Neutral	0	0	2	**138**	4	1	0	145	.786
	Total	129	59	95	167	53	7	70	**580**	—
7(M)	Angry	**95**	20	1	14	7	2	6	145	.616
	Fear	0	**84**	0	28	24	0	9	145	.363
	Happy	4	29	**56**	1	5	3	47	145	.379
	Neutral	2	1	0	**101**	40	0	1	145	.489
	Total	101	134	57	144	76	5	63	**580**	—
8(M)	Angry	**128**	10	0	3	0	3	1	145	.785
	Fear	2	**131**	0	0	3	3	6	145	.839
	Happy	12	0	**86**	0	3	13	31	145	.539
	Neutral	2	0	0	**132**	10	1	0	145	.890
	Total	144	141	86	135	16	20	38	**580**	—
9(M)	Angry	**119**	9	1	3	2	7	4	145	.794
	Fear	3	**133**	1	0	7	0	1	145	.830
	Happy	0	4	**81**	35	21	1	3	145	.492
	Neutral	1	1	9	**117**	16	1	0	145	.609
	Total	123	147	92	155	46	9	8	**580**	—
10(M)	Angry	**145**	0	0	0	0	0	0	145	1.00
	Fear	0	**61**	4	2	76	0	2	145	.347
	Happy	0	2	**83**	1	6	4	49	145	.540
	Neutral	0	11	1	**94**	39	0	0	145	.628
	Total	145	74	88	97	121	4	51	**580**	—
11(F)	Angry	**139**	0	0	0	0	6	0	145	.932
	Fear	1	**85**	0	4	39	3	13	145	.466
	Happy	2	17	**12**	5	9	4	96	145	.083
	Neutral	1	5	0	**70**	64	4	1	145	.428
	Total	143	107	12	79	112	17	110	**580**	—
12(F)	Angry	**136**	0	0	2	1	6	0	145	.580
	Fear	5	**61**	2	18	52	7	0	145	.407
	Happy	67	1	**14**	8	1	28	26	145	.080
	Neutral	12	1	1	**129**	1	1	0	145	.731
	Total	220	63	17	157	55	42	26	**580**	—
13(F)	Angry	**124**	10	3	4	1	0	3	145	.675
	Fear	7	**111**	0	0	27	0	0	145	.486
	Happy	26	53	**14**	8	3	2	39	145	.080
	Neutral	0	1	0	**142**	2	0	0	145	.903
	Total	157	175	17	154	33	2	42	**580**	—
14(F)	Angry	**132**	5	1	4	0	2	1	145	.715
	Fear	9	**82**	0	3	1	3	47	145	.464
	Happy	9	11	**19**	12	1	0	93	145	.119
	Neutral	18	2	1	**115**	2	6	1	145	.681
	Total	168	100	21	134	4	11	142	**580**	—
15(F)	Angry	**142**	1	0	0	0	2	0	145	.721
	Fear	12	**102**	0	2	5	13	11	145	.619
	Happy	33	12	**19**	5	0	21	55	145	.096
	Neutral	6	1	7	**124**	5	2	0	145	.809
	Total	193	116	26	131	10	38	66	**580**	—
16(F)	Angry	**128**	11	0	1	3	2	0	145	.743
	Fear	0	**33**	4	37	9	0	62	145	.055
	Happy	21	92	**3**	0	26	0	3	145	.009
	Neutral	3	0	0	**128**	11	1	1	144	.681
	Total	152	136	7	166	49	3	66	**579**	—
17(F)	Angry	**113**	13	0	6	4	6	3	145	.468
	Fear	0	**53**	3	2	6	0	81	145	.231
	Happy	73	16	**6**	3	30	13	4	145	.028
	Neutral	2	2	0	**118**	22	1	0	145	.744
	Total	188	84	9	129	62	20	88	**580**	—
18(F)	Angry	**139**	1	0	2	0	3	0	145	.938
	Fear	0	**41**	1	27	16	0	60	145	.227
	Happy	3	7	**31**	0	2	5	97	145	.201
	Neutral	0	2	1	**137**	4	1	0	145	.780
	Total	142	51	33	166	22	9	157	**580**	—
19(F)	Angry	**80**	5	11	13	1	10	25	145	.525
	Fear	0	**126**	3	1	11	0	4	145	.788
	Happy	2	1	**102**	0	0	1	39	145	.552
	Neutral	2	7	14	**85**	35	0	2	145	.503
	Total	84	139	130	99	47	11	70	**580**	—
20(F)	Angry	**141**	3	0	0	0	1	0	145	.632
	Fear	0	**82**	0	6	38	0	19	145	.365
	Happy	73	42	**3**	6	8	9	4	145	.016
	Neutral	3	0	1	**127**	12	2	0	145	.800
	Total	217	127	4	139	58	12	23	**580**	—
21(F)	Angry	**119**	0	0	2	0	17	7	145	.421
	Fear	0	**38**	5	20	78	1	3	145	.216
	Happy	90	8	**16**	5	0	19	7	145	.084
	Neutral	23	0	0	**107**	2	13	0	145	.589
	Total	232	46	21	134	80	50	17	**580**	—
22(F)	Angry	**140**	0	0	0	0	5	0	145	.834
	Fear	0	**69**	1	21	19	0	35	145	.426
	Happy	15	8	**50**	0	46	4	22	145	.332
	Neutral	7	0	1	**121**	13	3	0	145	.711
	Total	162	77	52	142	78	12	57	**580**	—

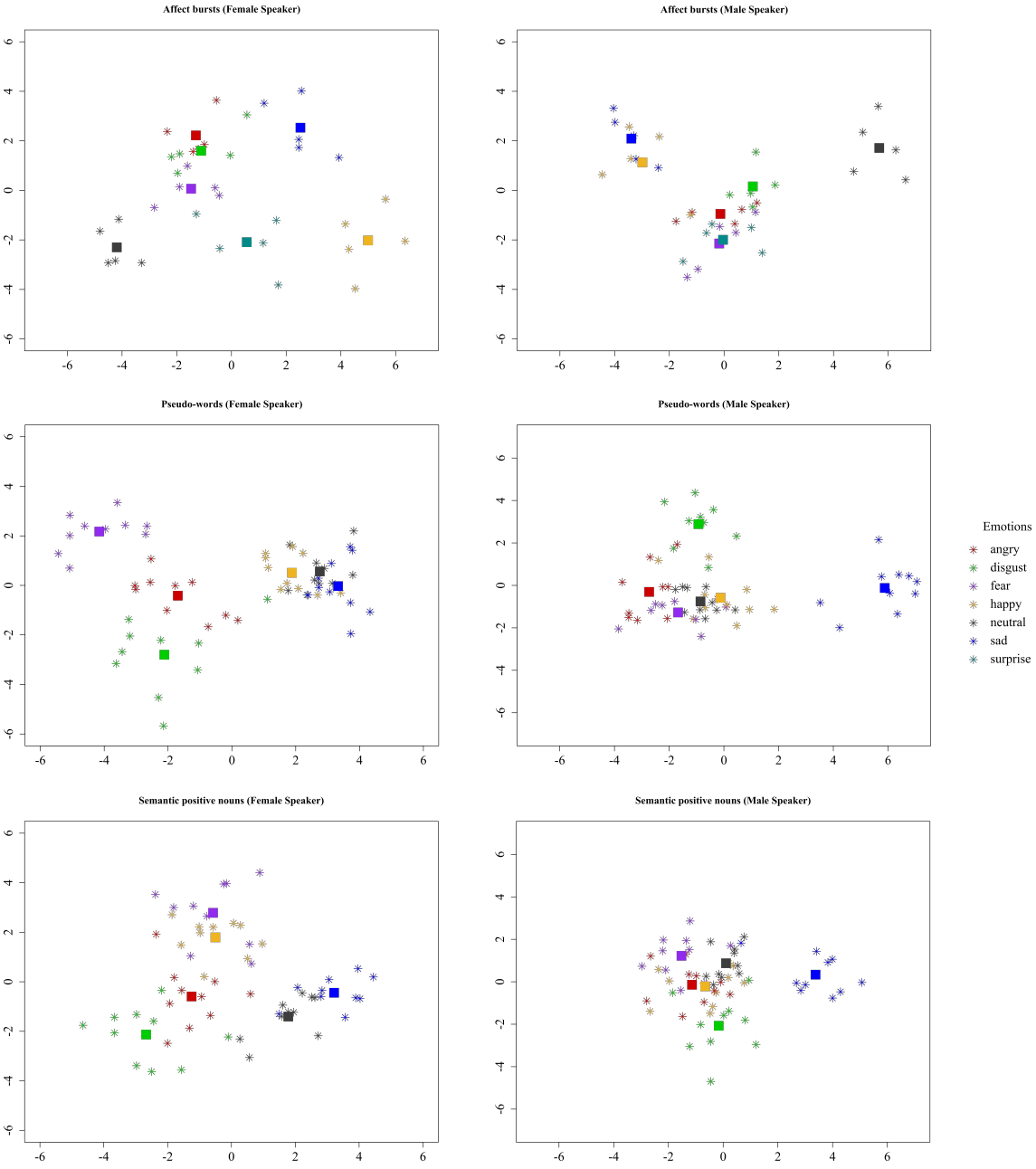
APPENDIX: STUDY 1

Linear discriminant analysis by gender and stimulus type



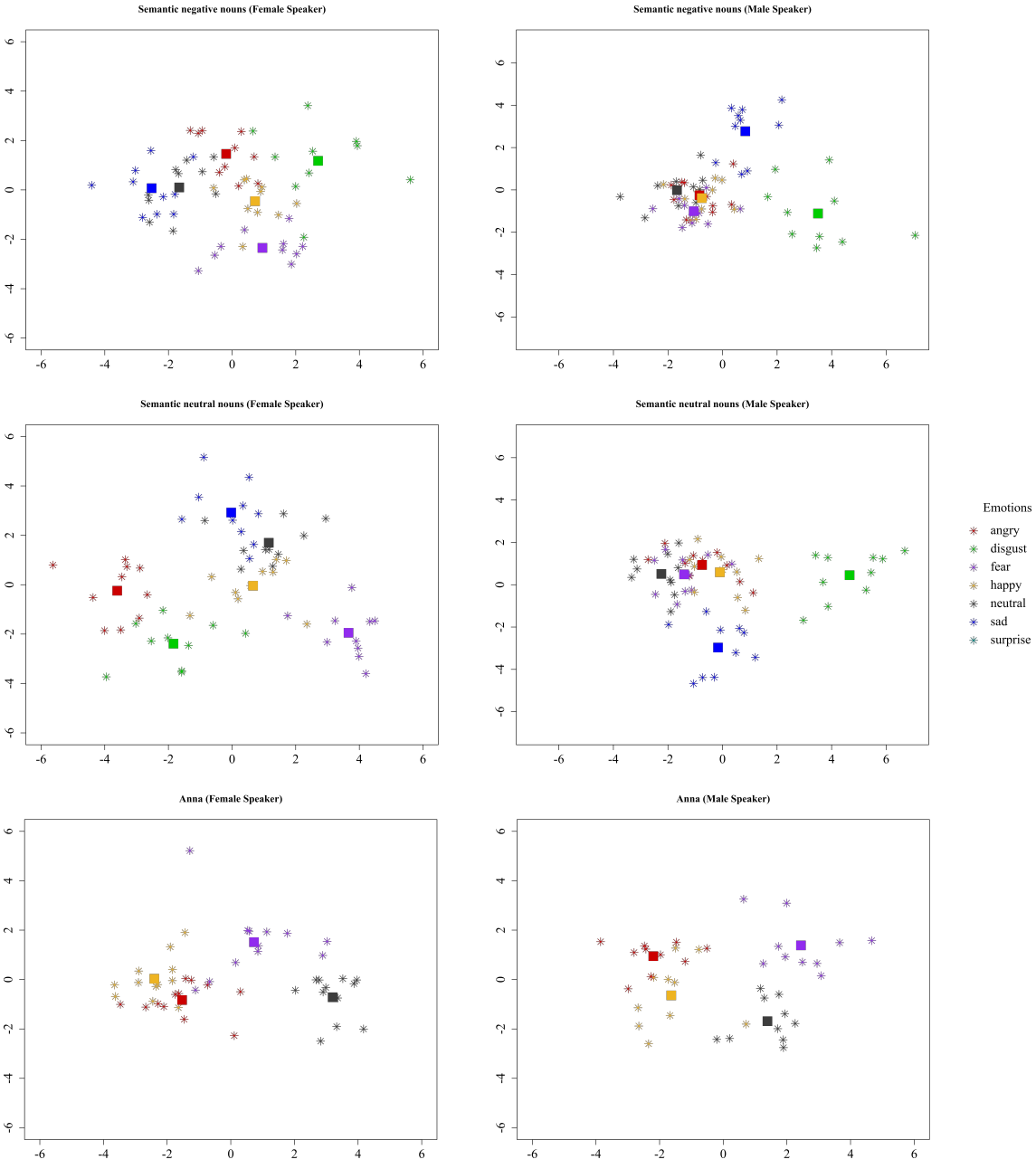
APPENDIX: STUDY 1

Linear discriminant analysis by gender and stimulus type



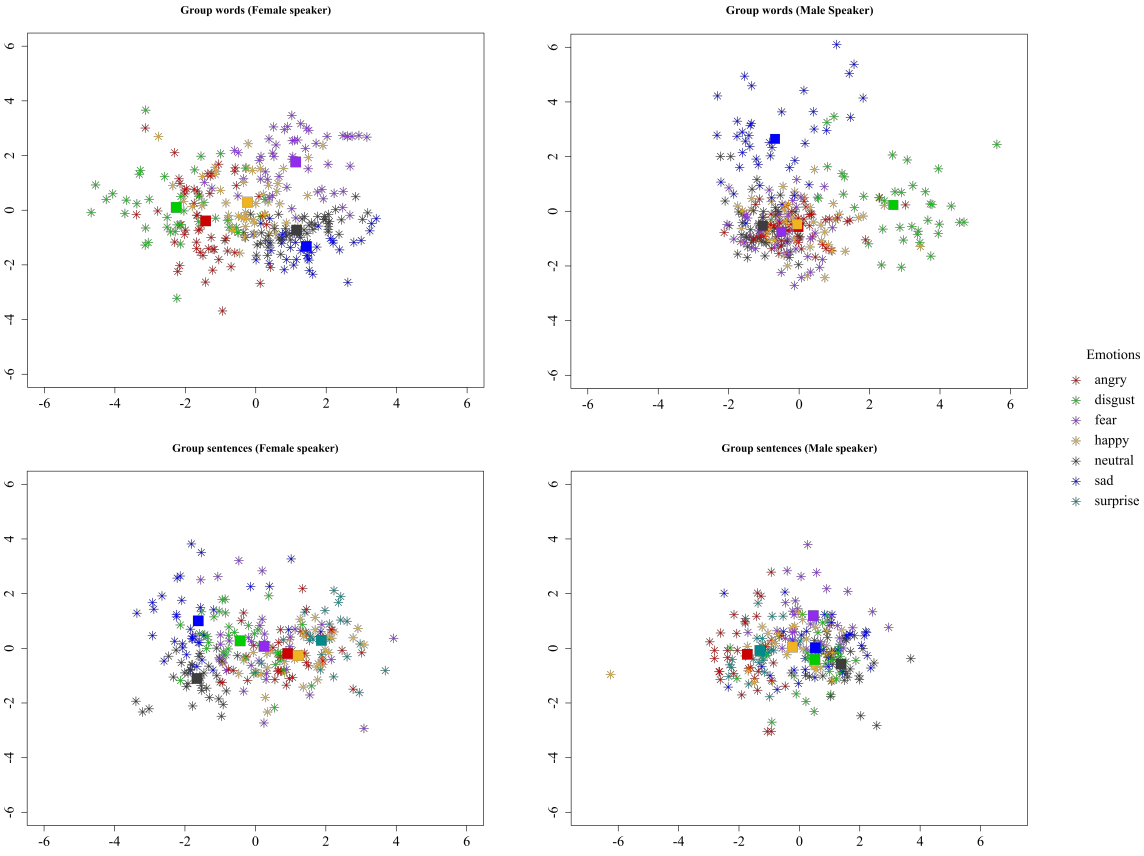
APPENDIX: STUDY 1

Linear discriminant analysis by gender and stimulus type



APPENDIX: STUDY 1

Linear discriminant analysis by gender and across all stimuli



APPENDIX: STUDY 2

(I) Additional analyses: Hormonal and modality specific effects on males' emotion recognition ability

Secondary analysis (H1, H2, H3)

A1 | H1: Overall performance accuracy in the three modalities

		Model terms	Df	Deviance	Resid.Df	Resid.Dev	Pr(>Chi)
Quasi-binomial logistic model (*DV* = ERA)	**H1a**	Null			121822	95628	
		Stimuli duration	1	137.60	121821	95491	< 0.001
		Emotions	5	2527.10	121816	95964	< 0.001
		Modalities	2	4431.90	121814	88532	< 0.001
		Emotions x Modalities	10	2135.20	121804	86397	< 0.001
	H1b	Null			81214	73621	
		Stimuli duration	1	109.97	81213	73511	< 0.001
		Emotions	5	2068.91	81208	71442	< 0.001
		Modalities	1	1462.93	81207	69980	< 0.001
		Emotions x Modalities	5	1982.71	81202	67997	< 0.001

Note: Model terms were inserted according to the analysis plan from pre-registration (osf.io/w2tgr). *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance; *DV* = dependent variable; *ERA* = emotion recognition accuracy.

A2 | H2: Emotion specificity and modality

	Model terms	Df	Deviance	Resid.Df	Resid.Dev	Pr(>Chi)
Quasi-binomial logistic model (H2a) (*DV* = ERA)	Null			60911	56686	
	Stimuli duration	1	154.58	60910	56532	< .001
	Emotions	2	580.61	60908	55951	< .001
	Modalities	2	3091.55	60906	52859	< .001
	Emotions x Modalities	4	112.67	60902	52747	< .001

		Df	SumSq	MeanSq	F-value	Pr(>F)
Linear model (H2a) (*DV* = RT)	Stimuli duration	1	263	263.30	105.611	< .001
	Emotions	2	281	140.60	56.459	< .001
	Modalities	2	3864	1931.78	775.714	< .001
	Emotions x Modalities	4	383	95.68	38.419	< .001
	Residuals	60902	151666	2.49		

	Model terms	Df	Deviance	Resid.Df	Resid.Dev	Pr(>Chi)
***Quasi-binomial logistic model* (H2b)** (*DV* = ERA)	Null			40607	25499	
	Stimuli duration	1	137.78	40606	25362	< .001
	Emotions	1	28.08	40605	25334	< .001
	Modalities	2	1531.07	40603	23802	< .001
	Emotions x Modalities	2	1651.83	40601	22151	< .001

		Df	SumSq	MeanSq	F-value	Pr(>F)
***Linear model* (H2b)** (*DV* = RT)	Stimuli duration	1	72	72.42	32.01	< .001
	Emotions	2	538	538.38	237.97	< .001
	Modalities	2	1103	551.27	243.66	< .001
	Emotions x Modalities	4	608	304.18	134.45	< .001
	Residuals	40601	91857	2.26		

Note: Model terms were inserted according to the analysis plan from pre-registration (osf.io/w2tgr). *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance; *DV* = dependent variable; *ERA* = emotion recognition accuracy; *RT* = reaction time.

A3 | H3: Testosterone

	Model terms	Df	Deviance	Resid.Df	Resid.Dev	Pr(>Chi)
***Quasi-binomial logistic model* (H3a)** (*DV* = ERA)	Null			60911	56686	
	Stimuli duration	1	154.58	60910	56532	< .001
	Emotions	2	580.61	60908	55951	< .001
	Modalities	2	3091.55	60906	52859	< .001
	Emotions x Modalities	4	112.67	60902	52747	< .001

		Df	SumSq	MeanSq	F-value	Pr(>F)
***Linear model* (H3b)** (*DV* = RT)	Stimuli duration	1	263	263.30	105.611	< .001
	Emotions	2	281	140.60	56.459	< .001
	Modalities	2	3864	1931.78	775.714	< .001
	Emotions x Modalities	4	383	95.68	38.419	< .001
	Residuals	60902	151666	2.49		

Note: Model terms were inserted according to the analysis plan from pre-registration (osf.io/w2tgr). *Resid. Df* = residual degrees of freedom; *Resid. Dev.* = Residual deviance; *DV* = dependent variable; *ERA* = emotion recognition accuracy; *RT* = reaction time.

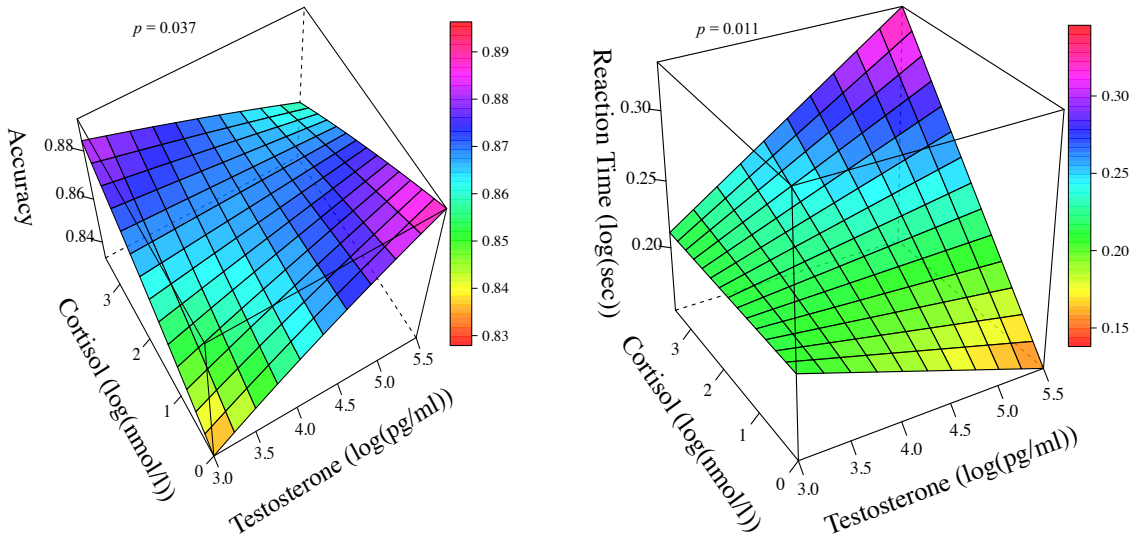


Figure A: Surface plots showing the TxC with ERA and RTs

Exploratory analyses for T, C and TxC terciles for each modality

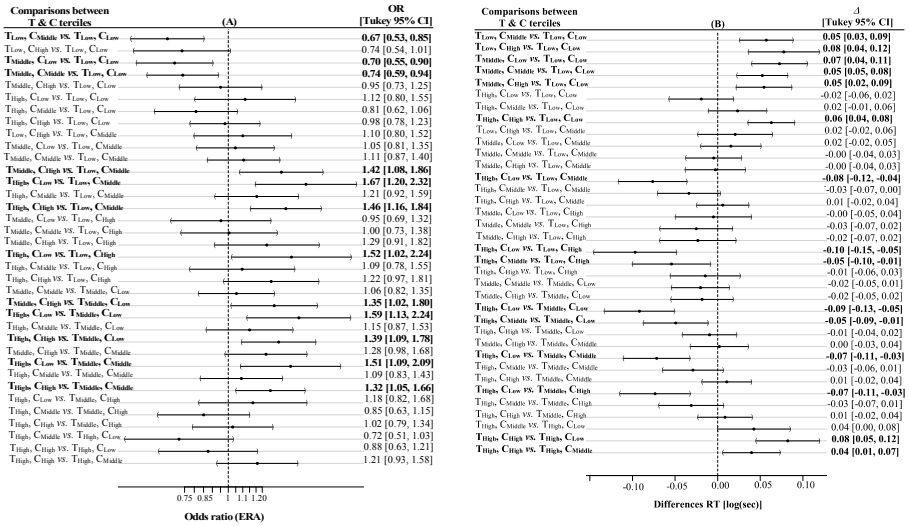
Table A4 | Logistic and linear models for T, C and TxC terciles

		Model terms	*Df*	*Deviance*	*Resid.Df*	*Resid.Dev*	*Pr(>Chi)*
Quasi-binomial logistic models (DV = *ERA*)	**Audio-visual**	Null			40607	19101	
		Stimuli duration	1	31.10	40606	19070	< 0.001
		Emotions	5	670.64	40601	18399	< 0.001
		Testosterone (T)	2	16.31	40599	18383	< 0.001
		Cortisol (C)	2	23.15	40597	18360	< 0.001
		TxC	4	26.41	40593	18334	< 0.001
	Audio	Null			40607	42539	
		Stimuli duration	1	55.85	40606	42484	< 0.001
		Emotions	5	2192.05	40601	40291	< 0.001
		Testosterone (T)	2	0.23	40599	40291	0.8922
		Cortisol (C)	2	0.24	40597	40291	0.8848
		TxC	4	16.93	40593	40274	0.0020
	Visual	Null			40606	29660	
		Stimuli duration	1	57.88	40605	29602	< 0.001
		Emotions	5	1910.38	40600	27692	< 0.001
		Testosterone (T)	2	2.18	40598	27690	0.3379
		Cortisol (C)	2	1.96	40596	27688	0.3771
		TxC	4	26.81	40592	27661	< 0.001

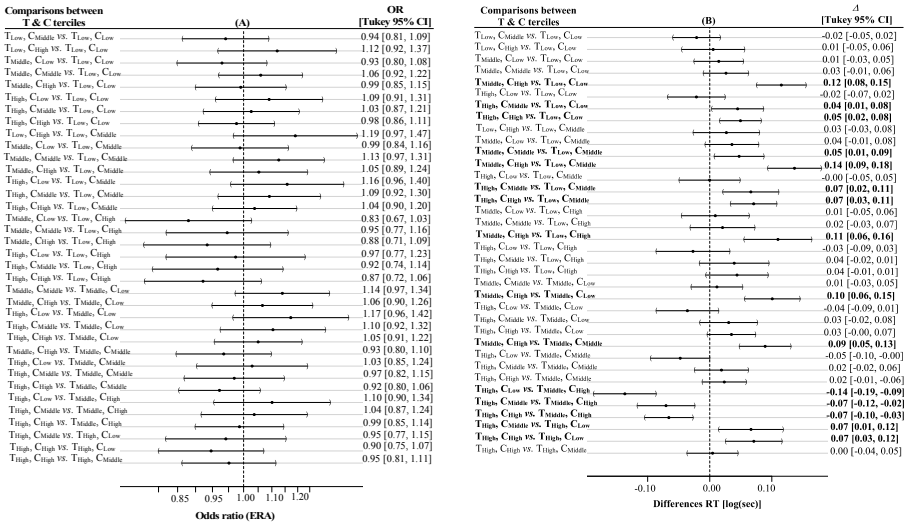
		Model terms	*Df*	*SumSq*	*MeanSq*	*F-value*	*Pr(>F)*
Linear models (DV = *RT*)	**Audio-visual**	Stimuli duration	1	37.80	37.76	134.89	< 0.001
		Emotions	5	284.30	56.85	203.13	< 0.001
		Testosterone (T)	2	5.70	2.89	10.21	< 0.001
		Cortisol (C)	2	14.10	7.03	25.13	< 0.001
		TxC	4	15.00	3.76	13.42	< 0.001
		Residuals	40593	11361.70	0.28		
	Audio	Stimuli duration	1	35.80	37.78	86.00	< 0.001
		Emotions	5	439.20	87.85	211.18	< 0.001
		Testosterone (T)	2	22.50	11.24	27.02	< 0.001
		Cortisol (C)	2	20.90	10.44	25.10	< 0.001
		TxC	4	16.30	4.07	9.79	< 0.001
		Residuals	40593	16885.50	0.42		
	Visual	Stimuli duration	1	160.30	160.25	651.63	< 0.001
		Emotions	5	379.10	72.83	308.34	< 0.001
		Testosterone (T)	2	12.80	6.41	26.08	< 0.001
		Cortisol (C)	2	6.60	3.29	13.36	< 0.001
		TxC	4	6.80	1.70	6.92	< 0.001
		Residuals	40592	9982.50	0.25		

Note: T and C were categorized using terciles. *Resid.Df* = residual degrees of freedom; *Resid.Dev* = residual deviance; *SumSq* = sum of squares; *DV* = dependent variable; *ERA* = emotion recognition accuracy; *RT* = reaction time.

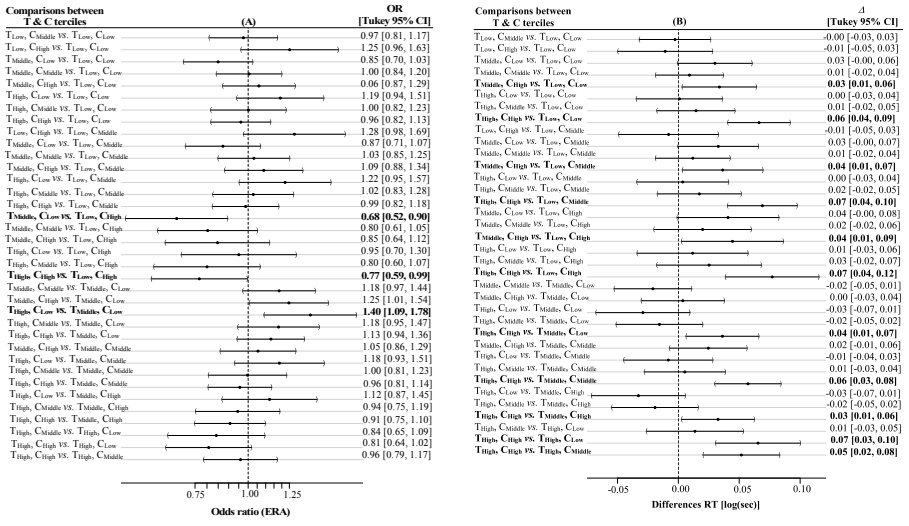
Audio-visual



Audio



Visual

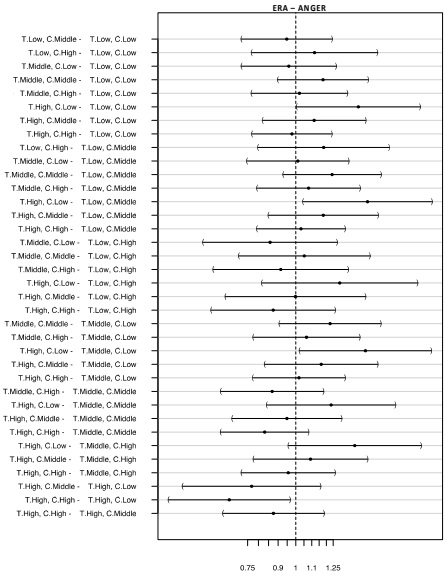


T, C and TxC terciles for each emotion category across all 3 modalities

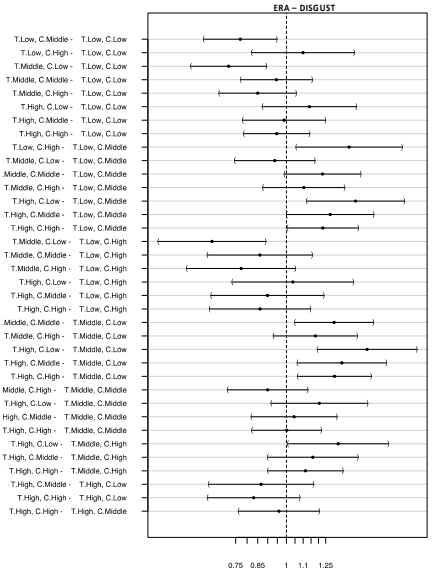
Table A5 | Logistic and linear models for T, C and TxC terciles for each emotion category across all three modalities

		Model terms	*Df*	*Deviance*	*Resid.Df*	*Resid.Dev*	*Pr(>Chi)*
Quasi-binomial logistic models (DV = *ERA*)	***Anger***	Null			20303	13569	
		Stimuli duration	1	77.38	20302	13492	< 0.001
		Modality	2	722.83	20300	12769	< 0.001
		Testosterone (T)	2	2.13	20298	12766	0.3450
		Cortisol (C)	2	2.72	20296	12764	0.2567
		TxC	4	15.98	20292	12748	0.0031**
	Disgust	Null			20303	21672	
		Stimuli duration	1	347.76	20302	21324	< 0.001
		Modality	2	1038.09	20300	20016	< 0.001
		Testosterone (T)	2	14.32	20298	20002	< 0.001
		Cortisol (C)	2	0.58	20296	20001	0.7508
		TxC	4	39.58	20292	19962	< 0.001
	Fear	Null			20303	15783	
		Stimuli duration	1	129.13	20302	15654	< 0.001
		Modality	2	1068.77	20300	14585	< 0.001
		Testosterone (T)	2	0.38	20298	14585	0.8313
		Cortisol (C)	2	4.40	20296	14580	0.1163
		TxC	4	11.46	20292	14569	0.0242*
	Happiness	Null			20303	11890.80	
		Stimuli duration	1	739.79	20302	11097.00	< 0.001
		Modality	2	2560.41	20300	8536.60	< 0.001
		Testosterone (T)	2	7.28	20298	8529.30	0.0725
		Cortisol (C)	2	3.32	20296	8526.00	0.3020
		TxC	4	11.35	20292	8514.60	0.0848
	Neutral	Null			20302	11450	
		Stimuli duration	1	45.57	20301	11404	< 0.001
		Modality	2	196.66	20299	11208	< 0.001
		Testosterone (T)	2	9.56	20297	11198	0.0083**
		Cortisol (C)	2	0.55	20295	11198	0.7586
		TxC	4	1.71	20291	11196	0.7881
	Sadness	Null			20303	18605	
		Stimuli duration	1	356.45	20302	18249	< 0.001
		Modality	2	868.76	20300	17380	< 0.001
		Testosterone (T)	2	0.60	20298	17380	0.7517
		Cortisol (C)	2	9.27	20296	17370	0.0117*
		TxC	4	13.08	20292	17357	0.0137*
		Model terms	***Df***	***SumSq***	***MeanSq***	***F-value***	***Pr(>F)***
Linear model (DV = *RT*)	***Anger***	Stimuli duration	1	6.60	6.586	20.022	< 0.001
		Modality	2	81.50	40.743	123.873	< 0.001
		Testosterone (T)	2	3.30	1.649	5.014	0.0067**
		Cortisol (C)	2	8.60	4.278	13.007	< 0.001
		TxC	4	11.40	2.860	8.696	< 0.001
		Residuals	20292	6674.30	0.329		
	Disgust	Stimuli duration	1	4.90	4.898	13.959	< 0.001
		Modality	2	416.80	208.406	593.899	< 0.001
		Testosterone (T)	2	4.30	2.174	6.194	0.0020**
		Cortisol (C)	2	8.20	4.111	11.714	< 0.001
		TxC	4	4.40	1.090	3.108	0.0144*
		Residuals	20292	7120.70	0.351		
	Fear	Stimuli duration	1	102.70	102.702	323.878	< 0.001
		Modality	2	285.20	142.619	449.758	< 0.001
		Testosterone (T)	2	5.70	2.846	8.975	< 0.001
		Cortisol (C)	2	8.30	4.174	13.162	< 0.001
		TxC	4	1.00	0.261	0.823	0.5101
		Residuals	20292	6434.60	0.317		
	Happiness	Stimuli duration	1	45.70	45.697	189.555	< 0.001
		Modality	2	305.40	152.699	633.412	< 0.001
		Testosterone (T)	2	7.10	3.537	14.672	< 0.001
		Cortisol (C)	2	5.90	2.950	12.236	< 0.001
		TxC	4	1.70	0.416	1.728	0.1408
		Residuals	20292	4891.90	0.241		
	Neutral	Stimuli duration	1	145.90	145.872	469.455	< 0.001
		Modality	2	158.20	79.078	254.496	< 0.001
		Testosterone (T)	2	10.00	4.984	16.040	< 0.001
		Cortisol (C)	2	3.90	1.954	6.288	0.0019**
		TxC	4	1.80	0.439	1.411	0.2273
		Residuals	20291	6304.90	0.311		
	Sadness	Stimuli duration	1	82.20	82.222	284.479	< 0.001
		Modality	2	86.70	43.333	130.954	< 0.001
		Testosterone (T)	2	4.50	2.265	6.844	0.0010**
		Cortisol (C)	2	4.00	2.020	6.106	0.0022**
		TxC	4	0.60	0.143	0.433	0.7852
		Residuals	20292	6714.60	0.331		

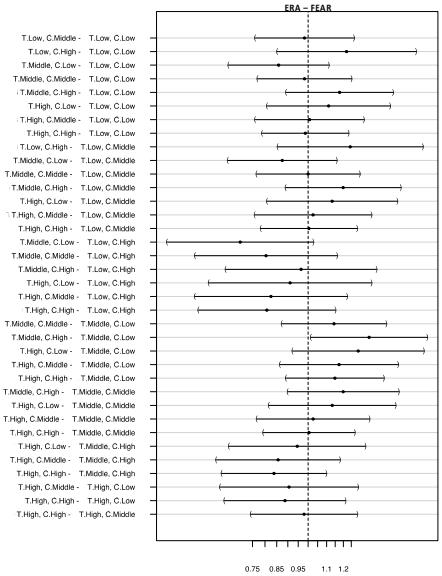
Note: *Resid.Df* = residual degrees of freedom; *Resid.Dev* = residual deviance; *SumSq* = sum of squares; *DV* = dependent variable; *ERA* = emotion recognition accuracy; *RT* = reaction time.



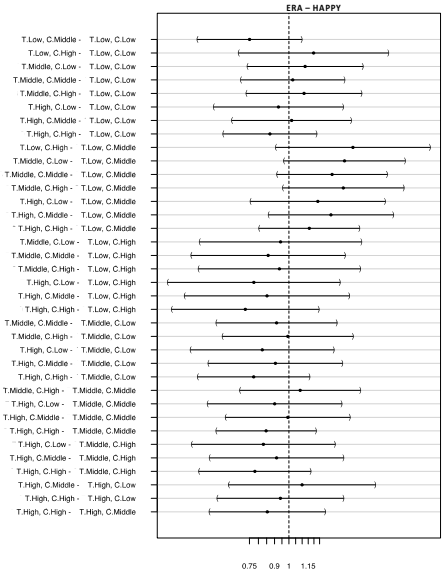
	OR	lwr	upr
T.Low, C.Middle - T.Low, C.Low	0.9471817	0.7226796	1.2414260
T.Low, C.High - T.Low, C.Low	1.1182502	0.7683347	1.6275244
T.Middle, C.Low - T.Low, C.Low	0.9588929	0.7229976	1.2717547
T.Middle, C.Middle - T.Low, C.Low	1.1769203	0.8988706	1.5409798
T.Middle, C.High - T.Low, C.Low	1.0220600	0.7674918	1.3610657
T.High, C.Low - T.Low, C.Low	1.4534520	1.0058237	2.1002913
T.High, C.Middle - T.Low, C.Low	1.1160873	0.8204408	1.5182703
T.High, C.High - T.Low, C.Low	0.9770462	0.7698865	1.2399481
T.Low, C.High - T.Low, C.Middle	1.1806079	0.7980411	1.7465704
T.Middle, C.Low - T.Low, C.Middle	1.0123642	0.7471791	1.3716674
T.Middle, C.Middle - T.Low, C.Middle	1.2425496	0.9280567	1.6636156
T.Middle, C.High - T.Low, C.Middle	1.0790538	0.7933853	1.4675809
T.High, C.Low - T.Low, C.Middle	1.5345018	1.0443923	2.2546084
T.High, C.Middle - T.Low, C.Middle	1.1783244	0.8492810	1.6348514
T.High, C.High - T.Low, C.Middle	1.0315298	0.7927752	1.3421885
T.Middle, C.Low - T.Low, C.High	0.8574940	0.5748524	1.2791039
T.Middle, C.Middle - T.Low, C.High	1.0524660	0.7119204	1.5559108
T.Middle, C.High - T.Low, C.High	0.9139815	0.6109551	1.3673054
T.High, C.Low - T.Low, C.High	1.2997556	0.8168805	2.0680685
T.High, C.Middle - T.Low, C.High	0.9980658	0.6569590	1.5162823
T.High, C.High - T.Low, C.High	0.8737277	0.6034581	1.2650423
T.Middle, C.Middle - T.Middle, C.Low	1.2273741	0.9066792	1.6614997
T.Middle, C.High - T.Middle, C.Low	1.0658751	0.7755310	1.4649185
T.High, C.Low - T.Middle, C.Low	1.5157606	1.0229870	2.2459036
T.High, C.Middle - T.Middle, C.Low	1.1639333	0.8306824	1.6308768
T.High, C.High - T.Middle, C.Low	1.0189315	0.7736270	1.3420180
T.Middle, C.High - T.Middle, C.Middle	0.8684191	0.6390806	1.1800571
T.High, C.Low - T.Middle, C.Middle	1.2349622	0.8411293	1.8131951
T.High, C.Middle - T.Middle, C.Middle	0.9483117	0.6840709	1.3146227
T.High, C.High - T.Middle, C.Middle	0.8301719	0.6386832	1.0790724
T.High, C.Low - T.Middle, C.High	1.4220809	0.9569523	2.1132861
T.High, C.Middle - T.Middle, C.High	1.0919978	0.7766849	1.5353190
T.High, C.High - T.Middle, C.High	0.9559577	0.7227893	1.2643452
T.High, C.Middle - T.High, C.Low	0.7678873	0.5087023	1.1591275
T.High, C.High - T.High, C.Low	0.6722246	0.4676698	0.9662499
T.High, C.High - T.High, C.Middle	0.8754209	0.6476361	1.1833215



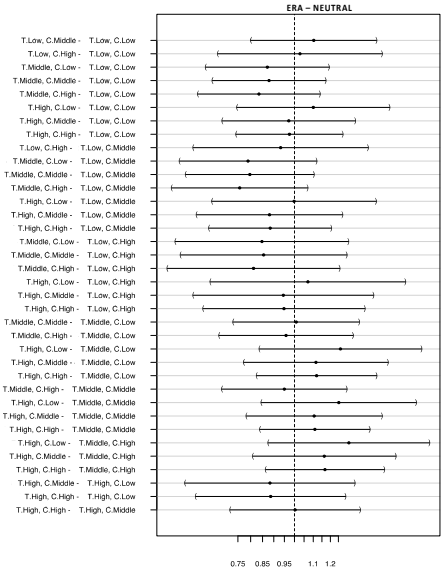
	OR	lwr	upr
T.Low, C.Middle - T.Low, C.Low	0.7697669	0.6256098	0.9471417
T.Low, C.High - T.Low, C.Low	1.0989422	0.8220202	1.4691538
T.Middle, C.Low - T.Low, C.Low	0.7207343	0.5819299	0.8926469
T.Middle, C.Middle - T.Low, C.Low	0.9450216	0.7716305	1.1573749
T.Middle, C.High - T.Low, C.Low	0.8497387	0.6830798	1.0570592
T.High, C.Low - T.Low, C.Low	1.1397389	0.8736434	1.4868820
T.High, C.Middle - T.Low, C.Low	0.9873812	0.7810778	1.2481749
T.High, C.High - T.Low, C.Low	0.9466329	0.7856402	1.1406161
T.Low, C.High - T.Low, C.Middle	1.4276299	1.0575945	1.9271347
T.Middle, C.Low - T.Low, C.Middle	0.9363020	0.7463065	1.1746668
T.Middle, C.Middle - T.Low, C.Middle	1.2276724	0.9888722	1.5241399
T.Middle, C.High - T.Low, C.Middle	1.1038909	0.8762133	1.3907288
T.High, C.Low - T.Low, C.Middle	1.4806286	1.1230422	1.9520736
T.High, C.Middle - T.Low, C.Middle	1.2827016	1.0027192	1.6408615
T.High, C.High - T.Low, C.Middle	1.2297657	1.0057136	1.5037319
T.Middle, C.Low - T.Low, C.High	0.6558437	0.4836346	0.8893717
T.Middle, C.Middle - T.Low, C.High	0.8599375	0.6390946	1.1570938
T.Middle, C.High - T.Low, C.High	0.7732332	0.5684382	1.0518111
T.High, C.Low - T.Low, C.High	1.0371236	0.7359337	1.4615790
T.High, C.Middle - T.Low, C.High	0.8984833	0.6529026	1.2364359
T.High, C.High - T.Low, C.High	0.8614037	0.6471857	1.1465277
T.Middle, C.Middle - T.Middle, C.Low	1.3111927	1.0495210	1.6381057
T.Middle, C.High - T.Middle, C.Low	1.1789901	0.9303199	1.4941288
T.High, C.Low - T.Middle, C.Low	1.5813579	1.1935107	2.0952411
T.High, C.Middle - T.Middle, C.Low	1.3699656	1.0650107	1.7622413
T.High, C.High - T.Middle, C.Low	1.3134284	1.0669042	1.6169158
T.Middle, C.High - T.Middle, C.Middle	0.8991738	0.7166858	1.1281283
T.High, C.Low - T.Middle, C.Middle	1.2060453	0.9179747	1.5845158
T.High, C.Middle - T.Middle, C.Middle	1.0448240	0.8199630	1.3313494
T.High, C.High - T.Middle, C.Middle	1.0017051	0.8231430	1.2190021
T.High, C.Low - T.Middle, C.High	1.3412817	1.0089362	1.7831023
T.High, C.Middle - T.Middle, C.High	1.1619823	0.8999460	1.5003154
T.High, C.High - T.Middle, C.High	1.1140283	0.9008413	1.3776667
T.High, C.Middle - T.High, C.Low	0.8663223	0.6435771	1.1661607
T.High, C.High - T.High, C.Low	0.8305700	0.6397325	1.0783359
T.High, C.High - T.High, C.Middle	0.9587309	0.7625870	1.2053248



	OR	lwr	upr
T.Low, C.Middle - T.Low, C.Low	0.9814343	0.7595354	1.268161
T.Low, C.High - T.Low, C.Low	1.2200280	0.8511960	1.748679
T.Middle, C.Low - T.Low, C.Low	0.8582768	0.6616744	1.113295
T.Middle, C.Middle - T.Low, C.Low	0.9812278	0.7687601	1.252417
T.Middle, C.High - T.Low, C.Low	1.1762827	0.8913752	1.552254
T.High, C.Low - T.Low, C.Low	1.1114369	0.8077934	1.529218
T.High, C.Middle - T.Low, C.Low	1.0068104	0.7594182	1.334794
T.High, C.High - T.Low, C.Low	0.9856689	0.7875568	1.233617
T.Low, C.High - T.Low, C.Middle	1.2431072	0.8531726	1.811258
T.Middle, C.Low - T.Low, C.Middle	0.8745127	0.6593347	1.159915
T.Middle, C.Middle - T.Low, C.Middle	0.9997896	0.7650103	1.306622
T.Middle, C.High - T.Low, C.Middle	1.1985343	0.8893597	1.615190
T.High, C.Low - T.Low, C.Middle	1.1324619	0.8080561	1.587105
T.High, C.Middle - T.Low, C.Middle	1.0258562	0.7579461	1.388464
T.High, C.High - T.Low, C.Middle	1.0043147	0.7822555	1.289410
T.Middle, C.Low - T.Low, C.High	0.7034894	0.4815481	1.027722
T.Middle, C.Middle - T.Low, C.High	0.8042666	0.5565611	1.162217
T.Middle, C.High - T.Low, C.High	0.9641440	0.6521109	1.425484
T.High, C.Low - T.Low, C.High	0.9109929	0.5975644	1.388818
T.High, C.Middle - T.Low, C.High	0.8252355	0.5563106	1.224161
T.High, C.High - T.Low, C.High	0.8079067	0.5662340	1.152727
T.Middle, C.Middle - T.Middle, C.Low	1.1432534	0.8715690	1.499627
T.Middle, C.High - T.Middle, C.Low	1.3705168	1.0136060	1.853103
T.High, C.Low - T.Middle, C.Low	1.2949633	0.9212988	1.820180
T.High, C.Middle - T.Middle, C.Low	1.1730603	0.8638830	1.592890
T.High, C.High - T.Middle, C.Low	1.1484277	0.8909860	1.480255
T.Middle, C.High - T.Middle, C.Middle	1.1987865	0.8989209	1.598683
T.High, C.Low - T.Middle, C.Middle	1.1327001	0.8157206	1.572854
T.High, C.Middle - T.Middle, C.Middle	1.0260720	0.7659759	1.374487
T.High, C.High - T.Middle, C.Middle	1.0045260	0.7923528	1.273514
T.High, C.Low - T.Middle, C.High	0.9448723	0.6633458	1.345880
T.High, C.Middle - T.Middle, C.High	0.8559256	0.6211050	1.179525
T.High, C.High - T.Middle, C.High	0.8379524	0.6387677	1.099248
T.High, C.Middle - T.High, C.Low	0.9058638	0.6336388	1.295043
T.High, C.High - T.High, C.Low	0.8868419	0.6478789	1.213944
T.High, C.High - T.High, C.Middle	0.9790015	0.7427602	1.290381

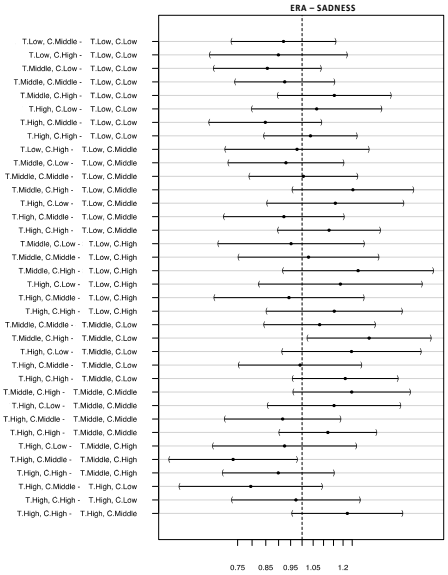


	OR	lwr	upr
T.Low, C.Middle - T.Low, C.Low	0.7504272	0.5131008	1.097525
T.Low, C.High - T.Low, C.Low	1.1964855	0.6933870	2.064615
T.Middle, C.Low - T.Low, C.Low	1.1254614	0.7392952	1.713339
T.Middle, C.Middle - T.Low, C.Low	1.0280566	0.7044645	1.500289
T.Middle, C.High - T.Low, C.Low	1.1158496	0.7334418	1.697640
T.High, C.Low - T.Low, C.Low	0.9260819	0.5778018	1.484294
T.High, C.Middle - T.Low, C.Low	1.0193220	0.6601732	1.573856
T.High, C.High - T.Low, C.Low	0.8702388	0.6192101	1.223035
T.Low, C.High - T.Low, C.Middle	1.5944058	0.9091896	2.796039
T.Middle, C.Low - T.Low, C.Middle	1.4997609	0.9649551	2.330972
T.Middle, C.Middle - T.Low, C.Middle	1.3699618	0.9175292	2.045488
T.Middle, C.High - T.Low, C.Middle	1.4869525	0.9572910	2.309672
T.High, C.Low - T.Low, C.Middle	1.2340729	0.7558815	2.014781
T.High, C.Middle - T.Low, C.Middle	1.3583223	0.8622822	2.139717
T.High, C.High - T.Low, C.Middle	1.1596579	0.8046809	1.671230
T.Middle, C.Low - T.Low, C.High	0.9406394	0.5216971	1.696008
T.Middle, C.Middle - T.Low, C.High	0.8592303	0.4907282	1.504451
T.Middle, C.High - T.Low, C.High	0.9326061	0.5174723	1.680774
T.High, C.Low - T.Low, C.High	0.7740018	0.4133603	1.449290
T.High, C.Middle - T.Low, C.High	0.8519301	0.4677101	1.551784
T.High, C.High - T.Low, C.High	0.7273292	0.4257635	1.242492
T.Middle, C.Middle - T.Middle, C.Low	0.9134535	0.5888687	1.416950
T.Middle, C.High - T.Middle, C.Low	0.9914597	0.6163684	1.594813
T.High, C.Low - T.Middle, C.Low	0.8228464	0.4882508	1.386739
T.High, C.Middle - T.Middle, C.Low	0.9056926	0.5557201	1.476065
T.High, C.High - T.Middle, C.Low	0.7732285	0.5146485	1.161729
T.Middle, C.High - T.Middle, C.Middle	1.0853971	0.7001346	1.682658
T.High, C.Low - T.Middle, C.Middle	0.9008083	0.5526633	1.468264
T.High, C.Middle - T.Middle, C.Middle	0.9915038	0.6305773	1.559016
T.High, C.High - T.Middle, C.Middle	0.8464892	0.5886463	1.217274
T.High, C.Low - T.Middle, C.High	0.8299343	0.4927060	1.397976
T.High, C.Middle - T.Middle, C.High	0.9134941	0.5608097	1.487976



	OR	lwr	upr
T.Low, C.Middle - T.Low, C.Low	1.1030730	0.8014748	1.518164
T.Low, C.High - T.Low, C.Low	1.0280718	0.6770852	1.561002
T.Middle, C.Low - T.Low, C.Low	0.8709709	0.6365210	1.191776
T.Middle, C.Middle - T.Low, C.Low	0.8787142	0.6586202	1.172358
T.Middle, C.High - T.Low, C.Low	0.8345567	0.6118441	1.138337
T.High, C.Low - T.Low, C.Low	1.1006536	0.7466828	1.622427
T.High, C.Middle - T.Low, C.Low	0.9714208	0.6925060	1.362672
T.High, C.High - T.Low, C.Low	0.9747663	0.7433023	1.278308
T.Low, C.High - T.Low, C.Middle	0.9320070	0.5974331	1.453949
T.Middle, C.Low - T.Low, C.Middle	0.7895859	0.5570726	1.119147
T.Middle, C.Middle - T.Low, C.Middle	0.7966056	0.5748361	1.103933
T.Middle, C.High - T.Low, C.Middle	0.7565743	0.5353029	1.069310
T.High, C.Low - T.Low, C.Middle	0.9978066	0.6575753	1.514074
T.High, C.Middle - T.Low, C.Middle	0.8806495	0.6074959	1.276624
T.High, C.High - T.Low, C.Middle	0.8836824	0.6473800	1.206238
T.Middle, C.Low - T.Low, C.High	0.8471888	0.5453239	1.316151
T.Middle, C.Middle - T.Low, C.High	0.8547206	0.5599527	1.304659
T.Middle, C.High - T.Low, C.High	0.8117689	0.5237013	1.258291
T.High, C.Low - T.Low, C.High	1.0705999	0.6517692	1.758573
T.High, C.Middle - T.Low, C.High	0.9448958	0.5973534	1.494640
T.High, C.High - T.Low, C.High	0.9481500	0.6283737	1.430659
T.Middle, C.Middle - T.Middle, C.Low	1.0088904	0.7321745	1.390187
T.Middle, C.High - T.Middle, C.Low	0.9581912	0.6815969	1.347028
T.High, C.Low - T.Middle, C.Low	1.2637087	0.8365102	1.909074
T.High, C.Middle - T.Middle, C.Low	1.1153308	0.7732315	1.608785
T.High, C.High - T.Middle, C.Low	1.1191720	0.8248055	1.518596
T.Middle, C.High - T.Middle, C.Middle	0.9497476	0.6913923	1.304644
T.High, C.Low - T.Middle, C.Middle	1.2525729	0.8449344	1.856876
T.High, C.Middle - T.Middle, C.Middle	1.1055025	0.7829895	1.560858
T.High, C.High - T.Middle, C.Middle	1.1093098	0.8391036	1.466527
T.High, C.Low - T.Middle, C.High	1.3188481	0.8751097	1.987591
T.High, C.Middle - T.Middle, C.High	1.1639961	0.8091584	1.674440
T.High, C.High - T.Middle, C.High	1.1680048	0.8636017	1.579704
T.High, C.Middle - T.High, C.Low	0.8825854	0.5731210	1.359149
T.High, C.High - T.High, C.Low	0.8856250	0.6048779	1.296678
T.High, C.High - T.High, C.Middle	1.0034439	0.7209093	1.396708

12



	OR	lwr	upr
T.Low, C.Middle - T.Low, C.Low	0.9202709	0.7292621	1.161309
T.Low, C.High - T.Low, C.Low	0.8994277	0.6621654	1.221704
T.Middle, C.Low - T.Low, C.Low	0.8561537	0.6740180	1.087507
T.Middle, C.Middle - T.Low, C.Low	0.9253513	0.7412394	1.155194
T.Middle, C.High - T.Low, C.Low	1.1541068	0.8964201	1.485869
T.High, C.Low - T.Low, C.Low	1.0669211	0.7988223	1.424999
T.High, C.Middle - T.Low, C.Low	0.8482769	0.6602590	1.089836
T.High, C.High - T.Low, C.Low	1.0376902	0.8433770	1.276773
T.Low, C.High - T.Low, C.Middle	0.9773510	0.7092558	1.346785
T.Middle, C.Low - T.Low, C.Middle	0.9303279	0.7192322	1.203380
T.Middle, C.Middle - T.Low, C.Middle	1.0055205	0.7899253	1.279958
T.Middle, C.High - T.Low, C.Middle	1.2540946	0.9574074	1.642721
T.High, C.Low - T.Low, C.Middle	1.1593554	0.8549440	1.572156
T.High, C.Middle - T.Low, C.Middle	0.9217687	0.7050933	1.205028
T.High, C.High - T.Low, C.Middle	1.1275921	0.8976502	1.416436
T.Middle, C.Low - T.Low, C.High	0.9518871	0.6874787	1.317989
T.Middle, C.Middle - T.Low, C.High	1.0288223	0.7524070	1.406786
T.Middle, C.High - T.Low, C.High	1.2831568	0.9174673	1.794605
T.High, C.Low - T.Low, C.High	1.1862222	0.8243617	1.706924
T.High, C.Middle - T.Low, C.High	0.9431296	0.6754189	1.316951
T.High, C.High - T.Low, C.High	1.1537227	0.8523240	1.561702
T.Middle, C.Middle - T.Middle, C.Low	1.0808238	0.8437294	1.384544
T.Middle, C.High - T.Middle, C.Low	1.3480136	1.0232819	1.775797
T.High, C.Low - T.Middle, C.Low	1.2461794	0.9143507	1.698433
T.High, C.Middle - T.Middle, C.Low	0.9907999	0.7535853	1.302685
T.High, C.High - T.Middle, C.Low	1.2120373	0.9584441	1.532728
T.Middle, C.High - T.Middle, C.Middle	1.2472093	0.9609875	1.618680
T.High, C.Low - T.Middle, C.Middle	1.1529903	0.8572131	1.550824
T.High, C.Middle - T.Middle, C.Middle	0.9167080	0.7077792	1.187310
T.High, C.High - T.Middle, C.Middle	1.1214013	0.9026075	1.393231
T.High, C.Low - T.Middle, C.High	0.9244562	0.6711967	1.273277
T.High, C.Middle - T.Middle, C.High	0.7350073	0.5524312	0.977924
T.High, C.High - T.Middle, C.High	0.8991284	0.7013197	1.152729
T.High, C.Middle - T.High, C.Low	0.7950700	0.5782067	1.093270
T.High, C.High - T.High, C.Low	0.9726025	0.7308803	1.294269
T.High, C.High - T.High, C.Middle	1.2232918	0.9561992	1.564991

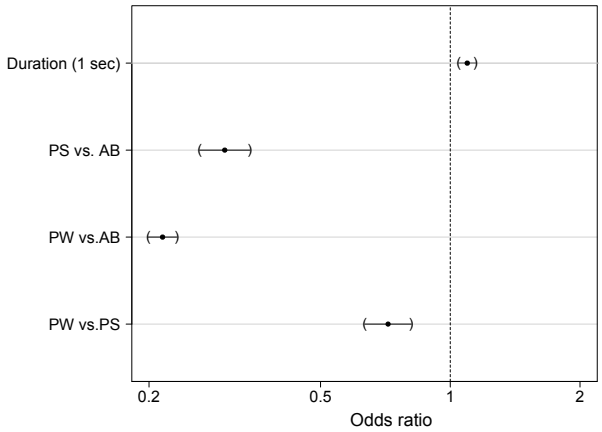
13

Hypothesis 3 (*Stimulus type/length*)

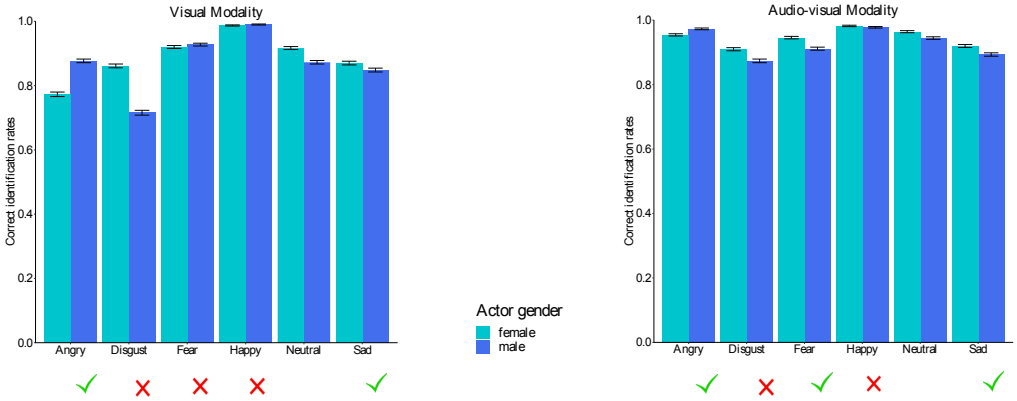
H3a: In the *audio* and *audio-visual* modalities, the performance accuracy would be influenced to a greater extent by the stimulus type (i.e., pseudo-speech, affect bursts) than by the length of the stimulus. ✓

H3b: In the *visual* modality, the recognition accuracy would be positively correlated with the length of the stimulus time-frame. ✗

OR = 0.938; [0.913; 0.963]



H4a: For the **visual** and **audio-visual** modality, we expected that *happy*, *sad* and *fearful* expressions would be better identified when displayed by an actress, whereas *angry* and *disgust* would have higher identification rates when displayed by an actor.



✓ ✗ ✗ ✗ ✓ ✓ ✗ ✓ ✗ ✓

Hypothesis 4 (*Gender of Encoder*)

H4b: For the audio modality we expect that recognition accuracy would not be systematically influenced by encoders' gender and the related stereotypes of emotional expressivity, but rather by the stimulus type carrying the vocal emotions. ✓ ($p < .0001$)

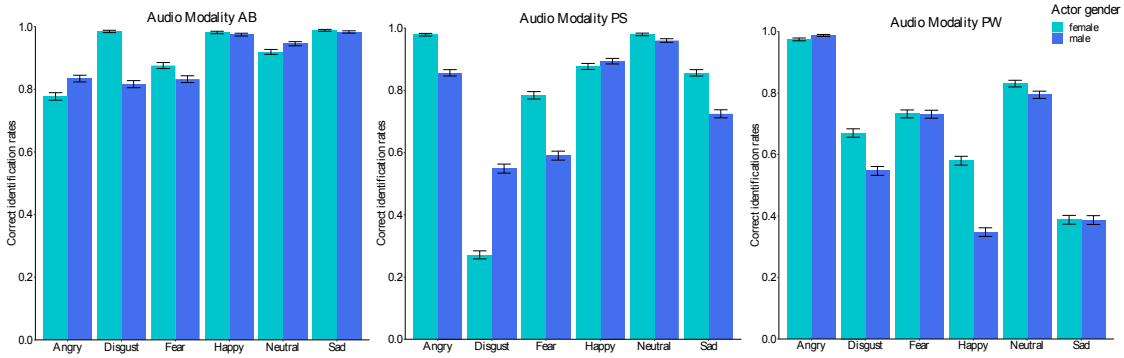


Table (A) | Quasi-binomial logistic model for ***emotion recognition*** with all predictors

Model terms	*Df*	*Deviance*	*Resid. Df*	*Resid. Dev*	*Pr(>Chi)*	*Direction*
NULL			126574	99227		
Duration	1	144.0	126573	99083	<2.2e-16***	-
Emotions	5	2663.8	126568	96419	<2.2e-16***	
Modality	2	4568.6	126566	91851	<2.2e-16***	
Testosterone T1	1	0.1	126565	91850	0.7047379	-
Testosterone T2	1	29.5	126564	91821	5.036e-08***	+
Cortisol T1	1	16.0	126563	91805	5.826e-05***	+
Cortisol T2	1	7.4	126562	91798	0.0064521**	***ns.***
MMG Affiliation (Hope of Affiliation)	1	0.4	126561	91797	0.5208092	***ns.***
MMG Affiliation (Fear of Rejection)	1	9.4	126560	91788	0.0021162**	-
MMG Achievement (Hope of Success)	1	54.6	126559	91733	1.223e-13***	+
MMG Achievement (Fear of Failure)	1	3.6	126558	91730	0.0556961.	+
MMG Power (Hope of Power)	1	1.5	126557	91728	0.2187267	-
MMG Power (Fear of losing Power)	1	81.1	126556	91647	<2.2e-16***	+
BFI Openness	1	33.4	126555	91614	6.474e-09***	+
BFI Conscientiousness	1	0.4	126554	91613	0.5429154	***ns.***
BFI Extraversion	1	20.4	126553	91593	5.762e-06***	-
BFI Agreeableness	1	102.1	126552	91491	<2.2e-16***	+
BFI Neuroticism	1	14.5	126551	91476	0.0001297***	***ns.***
MET Cognitive Empathy: positive stimuli	1	96.4	126550	91380	<2.2e-16***	+
MET Cognitive Empathy: negative stimuli	1	97.2	126549	91282	<2.2e-16***	+
MET Emotional Empathy: positive stimuli	1	7.3	126548	91275	0.0066926**	-
MET Emotional Empathy: negative stimuli	1	2.2	126547	91273	0.1390772	***ns.***
PANAS Positive Affect (T1)	1	3.3	126546	91270	0.0697081.	+
PANAS Negative Affect (T1)	1	0.2	126545	91270	0.6177338	+
PANAS Positive Affect (T2)	1	35.3	126544	91234	2.476e-09***	-
PANAS Negative Affect (T2)	1	43.5	126543	91191	3.582e-11***	-

Table (B) | Linear model for ***reaction times*** with all predictors

Model terms	*Df*	*SumSq*	*MeanSq*	*F-value*	*Pr(>F)*	*Direction*
NULL			126574	99227		
Duration	1	345	344.83	145.0831	<2.2e-16***	-
Emotions	5	2530	506.00	212.8936	<2.2e-16***	
Modality	2	5785	2892.35	1216.9172	<2.2e-16***	
Testosterone T1	1	1	1.30	0.5483	0.4590275	-
Testosterone T2	1	75	75.38	31.7135	1.791e-08***	+
Cortisol T1	1	218	218.33	91.8603	<2.2e-16***	+
Cortisol T2	1	7	7.02	2.9539	0.0856706.	+
MMG Affiliation (Hope of Affiliation)	1	8	8.37	3.5219	0.0605668.	***ns.***
MMG Affiliation (Fear of Rejection)	1	43	42.61	17.9295	2.294e-05***	+
MMG Achievement (Hope of Success)	1	2	2.50	1.0518	0.3050939	***ns.***
MMG Achievement (Fear of Failure)	1	32	32.19	13.5418	0.0002334***	+
MMG Power (Hope of Power)	1	1	1.42	0.5962	0.4400263	***ns.***
MMG Power (Fear of losing Power)	1	29	29.10	12.2426	0.0004673***	***ns.***
BFI Openness	1	39	39.48	16.6087	4.597e-05***	+
BFI Conscientiousness	1	55	55.14	23.1997	1.462e-06***	***ns.***
BFI Extraversion	1	55	55.21	23.2268	1.441e-06***	-
BFI Agreeableness	1	3	2.72	1.1446	0.2846867	+
BFI Neuroticism	1	92	92.22	38.8001	4.710e-10***	***ns.***
MET Cognitive Empathy: positive stimuli	1	20	19.54	8.2197	0.0041445**	+
MET Cognitive Empathy: negative stimuli	1	21	20.77	8.7384	0.0031164**	-
MET Emotional Empathy: positive stimuli	1	14	13.78	5.7977	0.0160481*	-
MET Emotional Empathy: negative stimuli	1	1	1.33	0.5604	0.4541144	***ns.***
PANAS Positive Affect (T1)	1	1	0.86	0.3629	0.5469034	***ns.***
PANAS Negative Affect (T1)	1	242	241.53	101.6204	<2.2e-16***	+
PANAS Positive Affect (T2)	1	5	5.06	2.1280	0.1446303	***ns.***
PANAS Negative Affect (T2)	1	245	244.64	102.9270	<2.2e-16***	+