J Nonverbal Behav (2017) 41:239–267
DOI 10.1007/s10919-017-0253-4

ORIGINAL PAPER

What is the Melody of That Voice? Probing Unbiased Recognition Accuracy with the Montreal Affective Voices

Margarida Vasconcelos[1] · Marcelo Dias[1,2] · Ana P. Soares[3] · Ana P. Pinheiro[1,4]

Published online: 6 April 2017

Abstract The present study aimed to clarify how listeners decode emotions from human nonverbal vocalizations, exploring unbiased recognition accuracy of vocal emotions selected from the Montreal Affective Voices (MAV) (Belin et al. in Trends Cognit Sci 8:129–135, 2008. doi:10.1016/j.tics.2004.01.008). The MAV battery includes 90 nonverbal vocalizations expressing anger, disgust, fear, pain, sadness, surprise, happiness, sensual pleasure, as well as neutral expressions, uttered by female and male actors. Using a forced-choice recognition task, 156 native speakers of Portuguese were asked to identify the emotion category underlying each MAV sound, and additionally to rate the valence, arousal and dominance of these sounds. The analysis focused on unbiased hit rates (H_u Score; Wagner in J Nonverbal Behav 17(1):3–28, 1993. doi:10.1007/BF00987006), as well as on the dimensional ratings for each discrete emotion. Further, we examined the relationship between categorical and dimensional ratings, as well as the effects of speaker's and listener's sex on these two types of assessment. Surprise vocalizations were associated with the poorest accuracy, whereas happy vocalizations were the most accurately recognized, contrary to previous studies. Happiness was associated with the highest valence and dominance ratings, whereas fear elicited the highest arousal ratings. Recognition accuracy and dimensional ratings of vocal expressions were dependent both on speaker's sex and

Electronic supplementary material The online version of this article (doi:10.1007/s10919-017-0253-4) contains supplementary material, which is available to authorized users.

Margarida Vasconcelos and Marcelo Dias have contributed equally to this work.

✉ Ana P. Pinheiro
ana.pinheiro@psi.uminho.pt

[1] Neuropsychophysiology Laboratory, CIPsi, School of Psychology, University of Minho, Braga, Portugal

[2] Faculty of Medicine, Institute of Molecular Medicine, University of Lisbon, Lisbon, Portugal

[3] Human Cognition Laboratory, CIPsi, School of Psychology, University of Minho, Braga, Portugal

[4] Faculty of Psychology, University of Lisbon, Lisbon, Portugal

listener's sex. Further, discrete vocal emotions were not consistently predicted by dimensional ratings. Using a large sample size, the present study provides, for the first time, unbiased recognition accuracy rates for a widely used battery of nonverbal vocalizations. The results demonstrated a dynamic interplay between listener's and speaker's variables (e.g., sex) in the recognition of emotion from nonverbal vocalizations. Further, they support the use of both categorical and dimensional accounts of emotion when probing how emotional meaning is decoded from nonverbal vocal cues.

Keywords Non-verbal vocalizations · Emotion · Unbiased accuracy · Valence · Arousal · Dominance

Introduction

The voice plays a critical role in human communication, plausibly representing the most important sound category in a social environment. Besides carrying speech information, the voice conveys important information about the speaker, such as his/her identity, sex, age, dominance, confidence, and, last but not the least, his/her emotional state (e.g., Belin et al. 2004; Jiang et al. 2015).

Vocal cues are among the most frequently used nonverbal cues to express emotions and to infer emotion from others (e.g., Laukka et al. 2005). Vocal emotions might be expressed via short bursts with no verbal content (nonverbal emotional vocalizations—Schröder 2003), or via the modulation of suprasegmental features of speech (emotional prosody—Paulmann and Kotz 2008). Nonverbal emotional vocalizations and emotional prosody rely on specific configurations of acoustic cues, such as pitch (F0), intensity, and duration (e.g., Juslin and Laukka 2003; Laukka et al. 2005), and are associated with different neurofunctional mechanisms (Fecteau et al. 2007; Liu et al. 2012; Morris et al. 1999; Sauter and Eimer 2010; Schirmer and Kotz 2003; Schirmer et al. 2004). In the context of social interactions, the accurate and fast recognition of vocal emotions is critical for detecting the intentions and predicting the behaviors of social partners, therefore influencing the success of social communication (e.g., Darwin 1998; Juslin and Laukka 2003).

There are two major accounts according to which vocal emotional stimuli may be characterized. A categorical account of emotion proposes that the affect system consists of a limited number of fundamental emotions (e.g., anger, sadness, surprise, happiness, fear, or disgust) that are triggered by specific conditions, elicit specific behavioral and biological patterns, and are considered to be more or less universal (e.g., Ekman 1992; Scherer and Ellgring 2007). A dimensional approach (Bradley et al. 2001b; Bradley and Lang 1994) proposes that vocal emotional stimuli may be defined based on specific dimensions, such as *valence* (a *continuum* ranging from unpleasant to pleasant), *arousal* (from calm to arousing), and *dominance* (from controlled to in control). The existing studies suggest that information provided by dimensions and categories is not redundant and should be combined to allow a deeper understanding of how emotional stimuli have an impact on perception, cognition, affect, and behavior (Kensinger and Corkin 2004; Kotz et al. 2013; Lewis et al. 2007; Stevenson and James 2008). Nevertheless, only a few studies adopted both perspectives in an attempt to characterize the processing of vocal emotional cues (Belin et al. 2008; Koeda et al. 2013; Lima et al. 2013, 2014; Sauter et al. 2010a; Sauter and Scott 2007; Stevenson et al. 2007).

In the last decades, a robust body of evidence has demonstrated consistent differences in the processing of neutral versus emotionally salient vocal stimuli (e.g., Laukka 2005; Liu et al. 2012; Paulmann and Kotz 2008; Pinheiro et al. 2013). Specifically, compared to neutral vocalizations, emotional vocalizations tend to be associated with lower (negative) or higher (positive) valence, and with increased arousal ratings (e.g., Belin et al. 2008); with increased bilateral amygdala activation (e.g., Fecteau et al. 2007); and with reduced N100 and increased P200 event-related potential (ERP) amplitude (e.g., Liu et al. 2012). Nonetheless, the study of emotion perception using vocal cues is much less well-developed than the study of emotion in faces. One of the reasons for the lack of studies probing how vocal emotions are perceived and recognized is plausibly related to the methodological challenge imposed by the control of dynamic stimuli such as the voice. Indeed, the processing of vocal stimuli relies on the combination of acoustic cues that change dynamically over time (e.g., Banse and Scherer 1996; Juslin and Laukka 2003), contrary to static faces whose low-level features remain unchanged during stimulus presentation. Therefore, the dynamic nature of vocal stimuli creates additional methodological challenges. Besides, in studies of vocal emotion, researchers often use prosodic speech samples differing in their acoustic structure (e.g., Paulmann and Kotz 2008; Pinheiro et al. 2013, 2014). However, the study of emotional cues in prosodic speech may be biased by the concurrent processing of semantic and syntactic information. The use of nonverbal vocalizations as experimental stimuli (e.g., laughs, cries, sighs, groans, screams) may therefore be advantageous (Kotchoubey et al. 2009; Sauter and Eimer 2010; see also Scherer et al. 1984). Compared to speech prosody, nonverbal vocalizations are not language-specific and, thereby, their use may facilitate the comparison of research results across different countries and/or cultures (Belin et al. 2008) by eliminating confounds associated with the simultaneous processing of verbal and vocal information. In that respect, they can be viewed as the auditory equivalent of facial emotional expressions, representing more primitive expressions of emotions (Belin et al. 2004). These features facilitate the study of vocal emotional processing not only in different cultures, but also in different age groups, as well as in patients with language deficits (e.g., Belin et al. 2008). Nevertheless, it is worth noting that probing both types of vocal emotional signals—speech prosody and vocalizations—provides crucial and complementary information on how humans decode the intentions and emotional states of social partners through their voice (Paulmann and Pell 2011).

Probing Nonverbal Emotional Vocalizations: Stimulus Sets

Sets of well-characterized and controlled nonverbal emotional vocalizations are scarce. Efforts to solve this limitation include the International Affective Digitized Sounds (IADS—Bradley and Lang 1999), and the Montreal Affective Voices (MAV—Belin et al. 2008) batteries. The IADS set consists of naturally occurring sounds (167 in the second version—Bradley and Lang 2007), associated with a wide number of contexts (e.g., human sounds—boy laugh, giggling; objects—music box, typewriter), and therefore not relying exclusively on voice stimuli. Each stimulus was rated in terms of valence, arousal, and dominance, based on a dimensional account of emotion. As an alternative, the MAV (Belin et al. 2008) was developed to provide a validated set of neutral and emotional nonverbal vocalizations. It consists of 90 nonverbal vocalizations expressing the emotions of anger, disgust, fear, pain, sadness, surprise, happiness, and pleasure, as well as neutral expressions, which were recorded by 5 male and 5 female actors, and rated by 29 Canadian participants. The original study of Belin et al. (2008) provided recognition accuracy scores, as well as ratings of valence, arousal, and intensity. More recently, a set of 121 nonverbal vocalizations expressing four positive emotions (achievement/triumph, amusement, pleasure, and relief) and four negative emotions (anger, disgust, fear, sadness) was developed by Lima et al. (2013), and rated by European Portuguese native speakers. However, neutral vocal sounds were not included in this stimulus set.

Despite the abovementioned efforts to create and validate emotional vocal stimuli, research probing emotion decoding from nonverbal vocalizations (reviewed below) is still underdeveloped. Considering the existing studies, methodological flaws preclude the generalization of the findings, such as small sample sizes or the absence of a neutral control condition. The inclusion of neutral vocal expressions represents an important methodological control, by allowing researchers to explore participants' responses to vocal stimuli representing a full range of valence assessments (from unpleasant to pleasant stimuli, with neutral stimuli in between), as well as to parametrically modulate valence ratings to understand how changes in the perceived pleasantness of a stimulus affect the behavioral or neural response of the listeners.

Probing Nonverbal Emotional Vocalizations: Recognition Accuracy

The studies that examined how emotions are inferred from nonverbal vocalizations demonstrated that the recognition rates for these stimuli are above chance levels (Table 1), in line with studies of emotional prosodic speech (Jiang et al. 2015; Pell et al. 2009a, b). However, the general mean recognition accuracy tends to be higher for nonverbal vocalizations than for emotional prosody (e.g., Banse and Scherer 1996).

Of note, caution is recommended in the interpretation and comparison of accuracy results from studies probing how vocal emotions are decoded, as chance levels may vary considerably across studies (e.g., Hall et al. 2008). For example, whereas participants were presented with ten response options in some studies (e.g., Sauter et al. 2010a), in others they were instructed to select between five response alternatives only (e.g., Sauter and Scott 2007). Because of differences in the number of response options, a given accuracy proportion does not mean the same when a common metric is lacking. In an attempt to normalize such differences, Rosenthal and Rubin (1989) proposed the *Proportion Index* (*pi* or π). This one-sample effect size estimator is based on the conversion of mean accuracy (proportion or percentage of correct responses, i.e., raw hits) to an equivalent proportion (or percentage) based on two options only, independently of the number of initial options provided to the participants (e.g., eight options in a forced-choice task). The index π ranges between 0 and 1, with 0.50 representing a chance-level response, and 1 representing 100% of accuracy. Therefore, it provides a more accurate metric for the comparison and interpretation of results from studies differing in their chance levels. In the following sections, when comparing accuracy levels across studies, we report both the original percentage of correct responses, as well as the *pi* (π) value (see Table 1).

The studies that examined how listeners decode vocal emotions have also consistently shown that recognition accuracy varies as a function of emotion type. In studies examining nonverbal vocalizations, *sadness* (97.16%/π = 1.00—Hawk et al. 2009; 79%/π = 0.97—Koeda et al. 2013), *disgust* (96.7%/π = 1.00—Lima et al. 2013; 93.5%/π = 0.99—Sauter et al. 2010a; 93.1%/π = 0.99—Schröder 2003; 74%/π = 0.95 in older adults—Lima et al. 2014), *relief* (70%/= 0.95—Laukka et al. 2013), *pleasure* (92.1%/π = 0.99 in younger

Table 1 Accuracy in decoding nonverbal emotional vocalizations

References	Speakers' Nationality	Participants' Nationality	Unbiased hit rates (Wagner 1993) reported	Accuracy analysis method	Mean recognition (*Chance level*)	Proportion Index* (π or PI—Rosenthal and Rubin 1989)
Belin et al. (2008)	Canadian	Canadian	No	Intensity derived accuracy scores	68.25% (*11.11%*)	0.95
Hawk et al. (2009)	Dutch	Dutch	Yes	Forced-choice hit rates	76.73% (*10.00%*)	0.97
Koeda et al. (2013)	Canadian	Japanese	No	Intensity derived accuracy scores	51.15% (*11.11%*)	0.89
Laukka et al. (2013)	India, Kenya, Singapore, and USA	Swedish	No	Forced-choice hit rates	44.11% (*11.11%*) for Indian speakers; 37.94% (*11.11%*) for Kenyan speakers; 41.56% (*11.11%*) for Singapore speakers; 44.44% (*11.11%*) for USA speakers	0.86—Indian speakers; 0.83—Kenyan speakers; 0.85—Singapore speakers; 0.87—USA speakers
Lima et al. (2014)	Portuguese British	Portuguese	Yes	Intensity derived accuracy scores	67.33% (*12.50%*)	0.94
Lima et al. (2013)	Portuguese	Portuguese	No	Forced-choice hit rates	86.00% (*12.50%*)	0.98
Sauter et al. (2010a, b)	British	British	No	Forced-choice hit rates	69.85% (*10.00%*)	0.95
Sauter et al. (2010b)	British Himba (Namibia)	British Himba (Namibia)	No	Forced-choice hit rates	94.75% (*50.00%*) for British speakers—British participants; 83,50% (*50.00%*) for Himba speakers—British participants; 64,50% (*50.00%*) for British speakers—Himba participants; 72,50% (*50.00%*) for Himba speakers—Himba participants	0.95 for British speakers—British participants; 0.84 for Himba speakers—British participants; 0.65 for British speakers—Himba participants; 0.73 for Himba speakers—Himba participants

Table 1 continued

References	Speakers' Nationality	Participants' Nationality	Unbiased hit rates (Wagner 1993) reported	Accuracy analysis method	Mean recognition (*Chance level*)	Proportion Index* (π or PI—Rosenthal and Rubin 1989)
Sauter and Scott (2007)	British	British Swedish	Yes	Forced-choice hit rates	75.34% (*20.00%*) for British participants; 68.80% (*20.00%*) for Swedish participants	0.92 for British participants; 0.90 for Swedish participants
Schröder (2003)	German	German	No	Forced-choice hit rates	81.10% (*10.00%*)	0.98

* Values range between 0 and 1. A value of $\pi = 0.5$ corresponds to a chance-level response. A value of $\pi = 1$ corresponds to 100% accuracy

adults—Lima et al. 2014), and *amusement* (90.4%/$\pi = 0.97$—Sauter and Scott 2007) were associated with the highest accuracy rates, whereas the recognition of *fear* (70%/$\pi = 0.94$—Lima et al. 2013; 25%/$\pi = 0.73$—Koeda et al. 2013), *guilt* (20%/$\pi = 0.67$—Laukka et al. 2013), *achievement* (65.3%/$\pi = 0.93$ in younger adults—Lima et al. 2014), *anger* (60.6%/$\pi = 0.93$—Schröder 2003; 39.5%/$\pi = 0.82$ in older adults—Lima et al. 2014), *pride* (30.11%/$\pi = 0.80$—Hawk et al. 2009), and *contentment* (52.4%/$\pi = 0.82$—Sauter and Scott 2007; 46%/$\pi = 0.89$—Sauter et al. 2010a) were the least accurate. Specifically, in the case of the MAV battery (Belin et al. 2008), *sadness* was the most accurately recognized emotion (86%/$\pi = 0.98$), whereas *pain* was associated with the lowest accuracy (51%/$\pi = 0.89$). Accuracy rates for the other discrete categories were: 56%/$\pi = 0.91$ (fear), 59%/$\pi = 0.92$ (pleasure), 60%/$\pi = 0.92$ (happiness), 75%/$\pi = 0.96$ (surprise), 78%/$\pi = 0.97$ (anger), and 81%/$\pi = 0.97$ (disgust).

The studies mentioned above have adopted different procedures to collect vocal portrayals of different emotions, as well as to probe how listeners judge the emotional properties of the vocal samples. Furthermore, recognition accuracy has been assessed directly via forced-choice categorization tasks (e.g., Lima et al. 2013, Laukka et al. 2013; Sauter et al. 2010a; Sauter and Scott 2007), or indirectly via ratings of intensity in different scales corresponding to a certain number of emotional conditions (e.g., MAV battery—Belin et al. 2008; Koeda et al. 2013; Lima et al. 2014). In the latter approach, the interpretation of recognition accuracy is constrained by the fact that it does not result from a selection of a specific emotion category in a standard forced-choice categorization task, but it is rather inferred from intensity ratings whose outcome represent an indirect measure of accuracy. As such, this methodological choice prevents comparison with most of the previous literature, in which forced-choice classification tasks were used (e.g., Hawk et al. 2009). Further, as mentioned before, it is worth noting that differences in the number of response alternatives (and associated chance levels) in categorization tasks may preclude the comparison of accuracy rates across studies in the absence of a common metric (Hall et al. 2008).

In addition, most of the studies aiming to probe vocal emotional recognition have examined raw hit scores (Lima et al. 2013; Sauter et al. 2010b; Schröder 2003)—which do not account for biases in listeners' responses (Wagner 1993), whereas only a few have probed unbiased hit scores (Hawk et al. 2009; Lima et al. 2014; Sauter and Scott 2007). Raw hit scores are based on the proportion of times the correct category is used to classify a certain type of stimulus, and do not consider the number of times the exact same category is erroneously used to label a different stimulus (e.g., 'happy' label for happy stimuli vs. 'happy' label for angry stimuli). As such, performance for a particular vocal category may be inflated by the disproportionate use of that category (e.g., Sauter et al. 2013). Therefore, unbiased hit rates represent a more accurate measure of emotional recognition.

Unbiased Hit Rates in Vocal Emotional Recognition

Unbiased hit rates, such as the H_u score (Wagner 1993), were proposed to control for biases related to both stimulus and listener. This accuracy measure was originally developed by Wagner (1993), and reflects "the joint probability that a stimulus category is correctly identified given that it is presented at all and that a response is correctly used given that it is used at all" (p. 3). The H_u score takes into consideration not only the number of hits, as well as the false alarms and the biases resulting from the inappropriate use of categories.

Thereby, this is the most precise measure of recognition accuracy used in studies of nonverbal behavior because it considers both stimulus discriminability and participant's judgment accuracy (Wagner 1993). Nonetheless, it is worth noting that only a few studies used such correction in the analysis of accuracy data resulting from forced-choice procedures (see, for example, Hawk et al. 2009; Liu and Pell 2012; Sauter and Scott 2007). In these studies, 'raw' hit scores were also presented to allow an analysis of differences emerging from the use of the two types of measures ('raw' vs. unbiased hit rates). For example, in the study of Liu and Pell (2012) with prosodic pseudospeech, 'raw' hit rates revealed that neutral expressions were associated with the highest accuracy (86%) and pleasant surprise with the lowest accuracy (56%), whereas the H_u scores revealed that fear was the most accurately recognized emotion. In the case of the studies with nonverbal vocalizations conducted by Sauter and Scott (2007) and Hawk et al. (2009), both hit rates and H_u scores showed similar results, even though there were differences in the magnitude of the effects when using one or the other measure.

To the best of our knowledge, the study from Sauter and Scott (2007) was the first to report emotional recognition accuracy based on H_u scores, using a forced-choice task to categorize nonverbal vocalizations. Nonetheless, the analysis has only included positive emotions. Therefore, a set of validated nonverbal emotional (positive and negative), as well as neutral vocalizations, assessed both categorically and dimensionally, and providing unbiased measures of recognition accuracy is critical in emotion research.

Probing Nonverbal Emotional Vocalizations: The Effects of Sex

Most of the studies using nonverbal vocalizations did not examine whether and how speaker's or listener's sex affects recognition accuracy (Laukka et al. 2013; Lima et al. 2013; Sauter and Scott 2007; Sauter et al. 2010a, b). The studies that probed the role of sex differences did not reveal consistent effects (Hawk et al. 2009; Lima et al. 2014; Sauter et al. 2013): some observed that the effects of speaker's sex were not systematic or readily interpretable (Hawk et al. 2009), whereas others failed to find effects of participant's sex in derived accuracy and/or unbiased hit rates (Lima et al. 2014; Sauter et al. 2013). The studies that tested recognition accuracy using the MAV sounds (Belin et al. 2008; Koeda et al. 2013) revealed that vocal emotions are more accurately decoded when uttered by female than by male speakers. Further, listener's sex also modulates the capacity to decode emotions from vocal cues: female listeners tend to be more accurate at recognizing vocal emotions than male listeners (Belin et al. 2008). Due to the inconsistent findings, the role of both speaker's and listener's sex needs to be more systematically examined in studies of vocal emotional recognition.

The Current Study and Hypotheses

This study aimed to clarify how listeners decode emotional cues conveyed through nonverbal vocalizations taken from a widely used stimulus battery—the MAV.

Using a large sample and a forced-choice procedure, we provide unbiased recognition accuracy rates for positive, negative, and neutral vocalizations. Furthermore, this study attempted to conciliate the two prevailing theoretical accounts of emotion by combining both categorical (recognition accuracy) and dimensional (valence, arousal, and dominance) ratings. Additionally, we examined the effects of sex (speaker's and listener's) on these ratings (Belin et al. 2008; Koeda et al. 2013). We also probed whether and how recognition of discrete emotions would be predicted by the affective dimensions under analysis.

Based on existing studies (Belin et al. 2008; Hawk et al. 2009; Koeda et al. 2013; Laukka et al. 2013; Lima et al. 2013, 2014; Pell et al. 2009a, b; Sauter et al. 2010a, b; Sauter and Scott 2007; Schröder 2003), we expected accuracy levels to be above chance and to vary as a function of the emotional category. Specifically, sadness was expected to achieve the highest recognition rates (Belin et al. 2008; Koeda et al. 2013).

Dimensional ratings were expected to differ based on the discrete emotion type: negative (anger, disgust, fear, pain, sadness) and positive (happiness, pleasure) categories were expected to have extreme and opposite valence ratings; vocalizations rated as more unpleasant and as more pleasant were expected to receive higher arousal ratings compared to sounds with intermediate valence ratings (Bradley and Lang 2007; Lang et al. 2008; Soares et al. 2013); and negative vocalizations were expected to receive lower dominance ratings than the other types of sounds (Bradley et al. 2001a; Soares et al. 2012; Pinheiro et al., in press). We also expected both categorical and dimensional ratings to differ as function of listener's and speaker's sex.

Regarding the relationship between categorical and dimensional ratings, we hypothesized that the recognition of specific vocal emotions would be predicted by specific affective dimensions, as previously shown for other stimulus types (e.g., affective sounds—Stevenson and James 2008; Soares et al. 2013; written words—Stevenson et al. 2007; Soares et al. 2012; written sentences—Pinheiro et al., in press).

Together, these analyses aimed to extend the work of Belin et al. (2008): first, by allowing a finer (unbiased) characterization of the recognition accuracy rates; second, by providing a more detailed analysis of the dimensional structure of the MAV sounds; third, by shedding light on the relationship between categorical and dimensional properties of nonverbal vocal emotions; fourth, by clarifying the effects of listener's sex based on a larger sample size.

Method

Participants

A total of 156 participants (90 women and 66 men), with ages between 17 and 45 years ($M = 21.00$; $SD = 3.62$) participated in the study. They were selected if their native language was European Portuguese, and if they reported normal audition and normal or corrected-to-normal visual acuity. Participants whose responses were illustrative of non-discriminative ratings or suggestive of inattention (e.g., using repeatedly the same values across dimensions to rate different sounds) were excluded from the total sample (10 were excluded and not considered for the analyses). The ethics committee of the University of Minho approved the study. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

Stimuli

Stimuli were 90 nonverbal emotional vocalizations that are part of the MAV (Belin et al. 2008). Emotional and neutral vocal expressions were initially portrayed by 22 actors. Based on the ratings of 29 participants, the vocalizations produced by 10 actors (5 women and 5 men) were selected. These include 10 vocalizations for each emotional category (anger, disgust, fear, pain, sadness, surprise, happiness, and sensual pleasure), as well as 10 neutral expressions. Each category comprised vocalizations uttered by five male and five female actors (Belin et al. 2008). The duration of the vocalizations ranged from 385 ms (ms) for surprise stimuli to 2229 ms for sad stimuli.[1]

Procedure

The experiment took place in a quiet room and was run in a laboratory setting, with groups of no more than 10 participants. Stimulus presentation and timing was controlled with SuperLab Pro 4.5 (Cedrus Corporation, San Pedro, California, USA).

After providing informed consent, participants were given a booklet that included the numerical codes for the MAV sounds. Participants were instructed to choose the Self-Assessment Manikin (SAM) scale (Bradley and Lang 1994) number that better represented the way they felt while listening to that sound, for the affective dimensions of valence, arousal and dominance. They were also instructed to choose the emotional category that better characterized the vocalization, among nine options: ‘surprise’, ‘sadness’, ‘pleasure’, ‘pain’, ‘neutral’, ‘happiness’, ‘fear’, ‘disgust’, and ‘anger’. The order in which the ratings were performed was not fixed. Participants rated the vocalizations in the dimensional and categorical assessments in the order they preferred. They were encouraged to register their responses as quickly as possible, following their first impressions. Any questions or doubts were answered before the beginning of the task. Participants were also asked to provide socio-demographic (e.g., sex, age, nationality, lateralization, auditory and visual acuity) and linguistic (e.g., native language, second language learned) information.

The task was performed in a paper-and-pencil format. Participants listened to each vocalization that was played back to them via loudspeakers, and rated its valence (ranging from 1—“unpleasant” to 9—“pleasant”), arousal (ranging from 1—“calm” to 9—“aroused”), and dominance (ranging from 1—“controlled” to 9—“in control”), using the 9-point SAM scale (Bradley and Lang 1994). Furthermore, participants also classified each stimulus on a forced-choice task, selecting one of 9 emotional categories: surprise, sadness, pleasure, pain, neutral, happiness, fear, disgust, and anger. Before the experiment, examples representing each type of vocalization were provided and participants were shown how to use each scale.

Each vocalization was presented only once. The presentation of the vocal sounds was pseudo-randomized to avoid the consecutive presentation of two vocalizations from the same category (neutral or emotional). Before listening to the vocalization, the sentence “Please rate the next sound in line number __” appeared in the center of the screen (5 s duration), instructing participants to find the line in the booklet that referred to the sound they would hear. After the presentation of the vocalization, participants were instructed to categorize the vocal sound and to rate it in the three affective dimensions. The total response time was 17 s. After this time period, the next trial automatically began. Each trial lasted approximately 26 s and the entire procedure took 45 min to complete. Participants listened to and rated all vocalizations included in the MAV battery. To minimize fatigue and distraction, the experiment was broken up into three different blocks, and the pauses between blocks lasted 1 min each.

[1] Differences in the duration of the MAV vocalizations could represent a confounding factor. However, in naturalistic contexts, vocal emotions rely on both acoustic and temporal differences to ensure they are accurately communicated. Therefore, the duration of the MAV stimuli was not manipulated to keep the sounds closer to vocal expressions typically found in real-life social communication contexts.

This procedure differed slightly from the strategy adopted in the original validation by Belin et al. (2008) to meet the main goals of the current study. The SAM scale (Bradley and Lang 1994) was used to collect affective ratings of valence, arousal, and dominance, as it is more typically used in dimensional assessments in emotion research (e.g., Pinheiro et al. 2016, in press; Soares et al. 2012, 2013, 2015). Also, an explicit emotional categorization task replaced the implicit categorization task with intensity scales that was adopted by Belin et al. (2008), with the aim of understanding which emotional label is more often unequivocally selected for a given vocal sound. Nonetheless, these differences do not preclude the comparison of the current findings with the results obtained in the studies of Belin et al. (2008) and Koeda et al. (2013).

Statistical Analyses

The IBM SPSS Statistics 22.00 (SPSS, Corp., USA) software package was used for statistical analyses. Only significant results are presented ($p < 0.05$).

Categorical Assessment

Uncorrected recognition rates ('raw' hit rates), unbiased recognition rates (H_u score; Wagner 1993), and confusion patterns were analyzed. Raw hit rates were calculated as the total percentage of times participants accurately decoded each MAV stimulus. Unbiased recognition rates were obtained by calculating the squared frequency of correct identifications of a given emotion (numerator), divided by the number of stimuli in that same emotional category, multiplied by the total number of times (across all scales) that the corresponding scale was used (denominator) (Wagner 1993). The H_u score varies between 0 and 1 (0 = no correct response; 1 = perfect performance). When no stimulus in a given category is correctly identified and the corresponding label is incorrectly used to classify the other vocal categories, H_u is equal to zero. When all stimuli in a given category are correctly identified and the corresponding label is not used to classify stimuli from the other categories, H_u is equal to 1. According to Wagner (1993), recognition accuracy will be higher than chance performance if, in addition to significant differences between H_u scores and chance proportions, we are able to confirm H_u values as higher than chance values. In order to test differences between the observed values of H_u and the corresponding values of chance proportions for each stimulus category, paired samples *t* tests were computed. The effects of MAV emotional category, speaker's sex, and listener's sex on H_u scores were assessed using a repeated-measures ANOVA: MAV emotional category and speaker's sex were included as within-subjects factors, and listener's sex was included as between-subjects factor. Confusion patterns were also calculated to obtain a broader understanding of the emotional categorization process (please see Supplementary Material). Main effects and interactions were followed up with pairwise comparisons using the Bonferroni correction for multiple comparisons. Effect sizes and significant effects are reported using the partial η-square method (η_p^2).

Dimensional Assessment

We tested the effects of emotional category, speaker's sex and listener's sex on the dimensional ratings (valence, arousal, and dominance) using repeated-measures ANOVA, with MAV emotional category and speaker's sex as within-subjects factors, and listener's sex as between-subjects factor. All the analyses of variance (ANOVAs) were corrected for non-sphericity using the Greenhouse–Geisser method (the original *df* is reported). Main effects and interactions were followed up with pairwise comparisons using the Bonferroni correction for multiple comparisons. Effect sizes and significant effects are reported using the partial η-square method (η_p^2).

Relationship Between Categorical and Dimensional Ratings

Multiple linear regressions were run to investigate whether dimensional ratings predict recognition accuracy performance for each category, following Stevenson et al. (2007). Valence, arousal and dominance ratings were taken as predictors, while H_u scores for each discrete vocal emotion were taken as dependent variables. Therefore, the three dimensional ratings were entered simultaneously, as independent variables, in multiple linear regressions performed individually for each emotional category (dependent variable).

Results

Recognition Accuracy

Raw Hit Rates

Table 2 illustrates the overall recognition accuracy (i.e., correct classification of a stimulus according to the a priori categorization) for each MAV category. The overall mean recognition accuracy rate was 62.83%/$\pi = 0.93$, ranging from 29.55%/$\pi = 0.77$ (angry) to 90.00%/$\pi = 0.99$ (happy), and the mean accuracy rate for emotional vocalizations only (excluding the neutral stimuli) was 59.79%/$\pi = 0.92$. All emotions were recognized with above-chance accuracy. This is confirmed by recognition values above raw chance levels (the chance level in a 9-alternative forced choice task is 11.11%, i.e., 1 out of 9), as well as by π values above 0.50 (Rosenthal and Rubin 1989). Table 2 shows that vocal expressions of happiness were the most accurately categorized (90.00%/$\pi = 0.99$), followed by neutral (87.12%/$\pi = 0.98$), sadness (80.13%/$\pi = 0.97$), pleasure (66.35%/$\pi = 0.94$), disgust (64.94%/$\pi = 0.94$), and pain (57.95%/$\pi = 0.92$) vocalizations. The emotions with recognition rates below 50% were fear (48.46%/$\pi = 0.88$), surprise (40.96%/$\pi = 0.85$), and anger (29.55%/$\pi = 0.77$). Even though recognized above the chance level, these vocal emotions were frequently misclassified (see Supplementary Table 1). Nonetheless, the percentage of categorizations was higher for the intended emotion than for all the non-intended ones.

H_u Scores

Table 3 illustrates the overall unbiased recognition rates for each MAV category. Following Wagner (1993), *t* tests confirmed that H_u scores were significantly different and higher than chance proportions for all emotional and neutral stimulus categories: surprise: $t(155) = -16.50$; neutral: $t(155) = -32.07$; disgust: $t(155) = -36.69$; anger: $t(155) = -18.13$; sadness: $t(155) = -41.31$; pleasure: $t(155) = -30.21$; happiness: $t(155) = -49.54$; fear: $t(155) = -21.76$; pain: $t(155) = -24.41$; $p < 0.001$ for all

Table 2 'Raw' hit rates (%) in the total sample, and in male and female listeners separately

MAV emotional category	Speaker ($n = 10$)	Listeners ($n = 156$)		
		All	Female	Male
Neutral	Female	81.80 (20.74)	82.44 (18.92)	80.91 (23.12)
	Male	92.44 (15.96)	92.89 (13.51)	91.82 (18.89)
	ALL	87.12 (15.74)	87.67 (13.33)	86.36 (18.62)
Happiness	Female	91.67 (16.69)	94.00 (12.16)	88.49 (21.07)
	Male	88.33 (17.74)	88.89 (15.61)	87.58 (20.39)
	ALL	90.00 (14.68)	91.44 (11.76)	88.03 (17.82)
Pleasure	Female	73.21 (23.45)	72.89 (21.27)	73.64 (26.29)
	Male	59.49 (27.45)	61.33 (26.78)	56.97 (28.34)
	ALL	66.35 (20.76)	67.11 (19.39)	65.30 (22.62)
Surprise	Female	40.00 (23.39)	36.67 (21.04)	44.55 (25.73)
	Male	41.92 (27.05)	35.78 (24.63)	50.30 (28.12)
	ALL	40.96 (21.96)	36.22 (19.00)	47.42 (24.14)
Anger	Female	19.10 (20.99)	20.44 (19.66)	17.27 (22.71)
	Male	40.00 (20.95)	44.00 (18.77)	34.55 (22.61)
	ALL	29.55 (17.20)	32.22 (16.20)	25.91 (17.97)
Sadness	Female	85.39 (22.18)	88.67 (18.25)	80.91 (26.12)
	Male	74.87 (25.18)	77.33 (23.84)	71.52 (26.73)
	ALL	80.13 (20.79)	83.00 (17.51)	76.21 (24.16)
Pain	Female	54.62 (24.29)	57.56 (22.80)	50.61 (25.83)
	Male	61.28 (28.44)	58.67 (28.09)	64.85 (28.73)
	ALL	57.95 (20.94)	58.11 (19.42)	57.73 (22.99)
Fear	Female	43.08 (21.93)	46.22 (20.26)	38.79 (23.50)
	Male	53.85 (28.34)	57.56 (24.87)	48.79 (31.99)
	ALL	48.46 (20.83)	51.89 (17.86)	43.79 (23.65)
Disgust	Female	78.85 (22.17)	80.67 (18.83)	76.36 (26.00)
	Male	51.03 (22.41)	51.56 (20.77)	50.30 (24.62)
	ALL	64.94 (17.87)	66.11 (14.59)	63.33 (21.58)

Mean and standard deviation [*M* (*SD*)] values are shown. The mean % of responses for a given emotional category, irrespective of speaker's sex, is shown in the row "ALL". The mean % of responses for a given emotional category, irrespective of listener's sex, is shown in the column "All"

comparisons. H_u scores ranged from 0.19 ($SD = 0.14$) for surprise to 0.78 ($SD = 0.19$) for happiness.

The ANOVA showed a significant main effect of emotional category on H_u scores, $F(8, 1232) = 322.32$, $p < 0.001$, $\eta_p^2 = 0.677$. Happy vocalizations were associated with the highest accuracy (0.78) compared to all other types of sounds ($p < 0.001$ for all comparisons), except sadness ($p > 0.05$). After happy vocalizations, vocal expressions of sadness were the most accurately classified (0.74), followed by neutral (0.59), disgust (0.58), pleasure (0.45), pain (0.29), fear (0.26), anger (0.21), and surprise (0.19). Surprise vocalizations were less accurately recognized than all the other types of vocal sounds ($p < 0.01$ for all comparisons), except anger. A significant main effect of speaker's sex

Table 3 Unbiased recognition rates (H_u scores; range = 0 to 1) in the total sample, and in male and female listeners separately

MAV emotional category	Speaker (n = 10)	Listeners (n = 156)		
		All	Female	Male
Neutral	Female	0.59 (0.26)	0.60 (0.24)	0.59 (0.28)
	Male	0.62 (0.25)	0.62 (0.24)	0.62 (0.27)
	ALL	0.59 (0.22)	0.59 (0.20)	0.59 (0.24)
Happiness	Female	0.83 (0.22)	0.86 (0.17)	0.79 (0.27)
	Male	0.75 (0.22)	0.78 (0.18)	0.71 (0.25)
	ALL	0.78 (0.19)	0.81 (0.15)	0.74 (0.23)
Pleasure	Female	0.52 (0.23)	0.54 (0.22)	0.49 (0.25)
	Male	0.48 (0.26)	0.49 (0.26)	0.47 (0.26)
	ALL	0.49 (0.20)	0.50 (0.19)	0.46 (0.20)
Surprise	Female	0.22 (0.15)	0.20 (0.13)	0.25 (0.18)
	Male	0.20 (0.17)	0.16 (0.16)	0.25 (0.18)
	ALL	0.19 (0.14)	0.16 (0.11)	0.23 (0.16)
Anger	Female	0.15 (0.18)	0.16 (0.18)	0.13 (0.18)
	Male	0.30 (0.19)	0.32 (0.16)	0.27 (0.21)
	ALL	0.21 (0.14)	0.23 (0.14)	0.18 (0.15)
Sadness	Female	0.81 (0.25)	0.85 (0.21)	0.75 (0.28)
	Male	0.69 (0.25)	0.72 (0.24)	0.65 (0.27)
	ALL	0.75 (0.22)	0.78 (0.19)	0.69 (0.26)
Pain	Female	0.25 (0.15)	0.27 (0.14)	0.22 (0.15)
	Male	0.37 (0.23)	0.36 (0.23)	0.40 (0.23)
	ALL	0.29 (0.14)	0.29 (0.14)	0.29 (0.15)
Fear	Female	0.22 (0.15)	0.23 (0.14)	0.20 (0.15)
	Male	0.34 (0.22)	0.34 (0.21)	0.34 (0.25)
	ALL	0.26 (0.15)	0.27 (0.13)	0.25 (0.16)
Disgust	Female	0.74 (0.24)	0.75 (0.22)	0.73 (0.27)
	Male	0.44 (0.23)	0.45 (0.21)	0.44 (0.25)
	ALL	0.59 (0.20)	0.59 (0.17)	0.57 (0.23)

Mean and standard deviation [*M* (*SD*)] values are shown. Mean H_u scores for a given emotional category, irrespective of speaker's sex, are shown in the row "ALL". Mean H_u scores for a given emotional category, irrespective of listener's sex, are shown in the column "All"

was also found, $F(1, 154) = 5.69$, $p = 0.018$, $\eta_p^2 = 0.036$, indicating that recognition accuracy was higher when vocalizations were produced by female speakers.

The ANOVA also revealed a significant interaction between emotional category and speaker's sex, $F(8, 1232) = 57.19$, $p < 0.001$, $\eta_p^2 = 0.271$: vocal expressions of disgust, sadness and happiness were better recognized when they were uttered by female speakers, whereas anger, fear and pain were better recognized when they were uttered by male speakers ($p < 0.001$ for all comparisons).

The ANOVA showed a significant interaction effect between emotional category and listeners' sex, $F(8, 1232) = 3.70$, $p < 0.001$, $\eta_p^2 = 0.023$. Pairwise comparisons indicated that female listeners were significantly more accurate in recognizing anger ($p = 0.019$) and sadness ($p = 0.012$) compared to male listeners, whereas male listeners were better at recognizing surprise ($p = 0.006$) than women.

Dimensional Assessment

Tables 4, 5 and 6 present mean ratings for valence, arousal and dominance for each MAV emotional category as a function of speaker's sex in the total sample, and in male and female listeners separately.

Table 4 Valence ratings for each MAV emotional category as a function of speaker's sex in the total sample, and in male and female listeners separately

MAV emotional category	Speaker ($n = 10$)	Listeners ($n = 156$)		
		All	Female	Male
Neutral	Female	4.91 (0.64)	4.88 (0.68)	4.94 (0.58)
	Male	4.89 (0.58)	4.86 (0.65)	4.93 (0.476)
	ALL	4.90 (0.56)	4.87 (0.60)	4.94 (0.49)
Happiness	Female	7.71 (1.02)	7.83 (1.04)	7.54 (0.97)
	Male	7.47 (0.97)	7.58 (0.94)	7.31 (1.01)
	ALL	7.59 (0.92)	7.70 (0.92)	7.43 (0.91)
Pleasure	Female	6.73 (1.16)	6.62 (1.11)	6.87 (1.21)
	Male	6.04 (1.18)	6.01 (1.22)	6.09 (1.13)
	ALL	6.39 (1.08)	6.32 (1.08)	6.48 (1.09)
Surprise	Female	3.95 (0.92)	3.87 (0.89)	4.05 (0.95)
	Male	4.08 (0.87)	3.96 (0.81)	4.25 (0.92)
	ALL	4.01 (0.80)	3.91 (0.75)	4.15 (0.87)
Anger	Female	3.60 (0.96)	3.43 (0.83)	3.83 (1.08)
	Male	4.11 (0.83)	3.96 (0.71)	4.30 (0.94)
	ALL	3.85 (0.80)	3.70 (0.67)	4.07 (0.91)
Sadness	Female	2.26 (1.12)	2.02 (0.91)	2.58 (1.29)
	Male	2.91 (1.29)	2.63 (1.21)	3.30 (1.31)
	ALL	2.59 (1.10)	2.33 (0.95)	2.94 (1.19)
Pain	Female	4.05 (1.15)	3.83 (1.10)	4.36 (1.14)
	Male	3.47 (0.95)	3.35 (0.92)	3.63 (0.98)
	ALL	3.76 (0.88)	3.59 (0.82)	3.99 (0.92)
Fear	Female	3.41 (1.03)	3.22 (0.99)	3.66 (1.04)
	Male	3.32 (0.94)	3.13 (0.86)	3.58 (0.99)
	ALL	3.36 (0.88)	3.17 (0.81)	3.62 (0.91)
Disgust	Female	3.61 (0.98)	3.45 (0.92)	3.83 (1.03)
	Male	4.04 (0.85)	3.94 (0.87)	4.17 (0.82)
	ALL	3.83 (0.80)	3.69 (0.76)	4.01 (0.83)

Mean and standard deviation [M (*SD*)] values are shown. The mean % of responses for a given emotional category, irrespective of speaker's sex, is shown in the row "ALL". The mean % of responses for a given emotional category, irrespective of listener's sex, is shown in the column "All"

Table 5 Arousal ratings for each MAV emotional category as a function of speaker's sex in the total sample, and in male and female listeners separately

MAV emotional category	Speaker ($n = 10$)	Listeners ($n = 156$)		
		All	Female	Male
Neutral	Female	3.40 (1.53)	3.61 (1.57)	3.13 (1.45)
	Male	3.20 (1.68)	3.42 (1.69)	2.91 (1.62)
	ALL	3.30 (1.55)	3.51 (1.57)	3.02 (1.49)
Happiness	Female	5.29 (2.12)	5.11 (2.20)	5.54 (1.99)
	Male	5.25 (2.04)	5.16 (2.11)	5.37 (1.95)
	ALL	5.27 (2.03)	5.14 (2.12)	5.45 (1.91)
Pleasure	Female	5.49 (1.80)	5.34 (1.73)	5.69 (1.89)
	Male	4.61 (1.82)	4.60 (1.67)	4.63 (2.02)
	ALL	5.05 (1.74)	4.97 (1.63)	5.16 (1.89)
Surprise	Female	4.86 (1.73)	5.10 (1.65)	4.54 (1.80)
	Male	4.61 (1.57)	4.76 (1.47)	4.40 (1.69)
	ALL	4.73 (1.58)	4.93 (1.48)	4.47 (1.68)
Anger	Female	5.14 (1.68)	5.34 (1.61)	4.86 (1.73)
	Male	5.10 (1.72)	5.29 (1.61)	4.84 (1.84)
	ALL	5.12 (1.62)	5.32 (1.52)	4.85 (1.71)
Sadness	Female	5.31 (1.76)	5.27 (1.64)	5.36 (1.92)
	Male	5.12 (1.77)	5.20 (1.61)	5.01 (1.96)
	ALL	5.21 (1.69)	5.23 (1.58)	5.19 (1.85)
Pain	Female	5.28 (1.76)	5.37 (1.62)	5.16 (1.94)
	Male	4.78 (1.64)	4.95 (1.54)	4.56 (1.76)
	ALL	5.03 (1.62)	5.16 (1.49)	4.85 (1.78)
Fear	Female	5.69 (1.81)	5.83 (1.66)	5.50 (1.99)
	Male	5.17 (1.78)	5.35 (1.58)	4.91 (2.01)
	ALL	5.43 (1.73)	5.59 (1.55)	5.21 (1.94)
Disgust	Female	4.65 (1.55)	4.84 (1.47)	4.39 (1.64)
	Male	4.23 (1.57)	4.41 (1.51)	3.98 (1.63)
	ALL	4.44 (1.50)	4.62 (1.43)	4.19 (1.56)

Mean and standard deviation [M (SD)] values are shown. The mean % of responses for a given emotional category, irrespective of speaker's sex, is shown in the row "ALL". The mean % of responses for a given emotional category, irrespective of listener's sex, is shown in the column "All"

Valence Ratings

The effect of MAV emotional category was significant, $F(8, 1240) = 637.95$, $p < 0.001$, $\eta_p^2 = 0.805$; see Table 4. Vocal expressions of happiness received the highest valence ratings ($M = 7.59$), followed by pleasure ($M = 6.39$), neutrality ($M = 4.90$), surprise ($M = 4.01$), anger ($M = 3.85$), disgust ($M = 3.83$), pain ($M = 3.76$), fear ($M = 3.36$), and sadness ($M = 2.59$). Even though planned pairwise comparisons revealed valence-related differences for all emotion types, valence ratings for angry vocalizations did not differ from surprise ($p = 0.097$), pain ($p = 1.000$), or disgust ($p = 1.000$), whereas

Table 6 Dominance ratings for each MAV emotional category as a function of speaker's sex in the total sample, and in male and female listeners separately

MAV emotional category	Speaker ($n = 10$)	Listeners ($n = 156$)		
		All	Female	Male
Neutral	Female	5.28 (1.98)	5.07 (1.92)	5.57 (2.03)
	Male	5.28 (2.23)	4.99 (2.22)	5.69 (2.21)
	ALL	5.28 (2.07)	5.03 (2.02)	5.63 (2.09)
Happiness	Female	6.50 (1.90)	6.57 (1.91)	6.41 (1.91)
	Male	6.39 (1.75)	6.46 (1.69)	6.29 (1.82)
	ALL	6.45 (1.78)	6.52 (1.76)	6.36 (1.81)
Pleasure	Female	5.95 (1.76)	5.94 (1.75)	5.97 (1.78)
	Male	5.49 (1.84)	5.37 (1.75)	5.65 (1.96)
	ALL	5.72 (1.71)	5.65 (1.69)	5.81 (1.74)
Surprise	Female	4.59 (1.84)	4.59 (1.73)	4.59 (1.99)
	Male	4.70 (1.76)	4.65 (1.61)	4.76 (1.97)
	ALL	4.65 (1.73)	4.63 (1.57)	4.68 (1.94)
Anger	Female	4.81 (1.73)	4.83 (1.64)	4.79 (1.86)
	Male	5.17 (1.76)	5.35 (1.62)	4.93 (1.92)
	ALL	4.99 (1.64)	5.09 (1.51)	4.86 (1.81)
Sadness	Female	4.22 (1.89)	4.30 (1.80)	4.12 (2.02)
	Male	4.59 (1.87)	4.53 (1.79)	4.67 (1.98)
	ALL	4.41 (1.82)	4.42 (1.73)	4.40 (1.95)
Pain	Female	4.80 (1.77)	4.86 (1.71)	4.73 (1.87)
	Male	4.64 (1.72)	4.59 (1.57)	4.71 (1.91)
	ALL	4.72 (1.65)	4.73 (1.51)	4.72 (1.83)
Fear	Female	4.58 (1.94)	4.73 (1.91)	4.38 (1.98)
	Male	4.46 (1.86)	4.53 (1.71)	4.35 (2.04)
	ALL	4.52 (1.82)	4.63 (1.73)	4.37 (1.94)
Disgust	Female	4.78 (1.75)	4.71 (1.61)	4.88 (1.93)
	Male	4.92 (1.78)	4.85 (1.70)	5.01 (1.89)
	ALL	4.85 (1.69)	4.78 (1.58)	4.95 (1.85)

Note. Mean and standard deviation [M (SD)] values are shown. The mean % of responses for a given emotional category, irrespective of speaker's sex, is shown in the row "ALL". The mean % of responses for a given emotional category, irrespective of listener's sex, is shown in the column "All"

valence ratings for disgust were not statistically different from surprise ($p = 0.066$), and pain ($p = 1.000$).

A significant interaction between MAV emotional category and speaker's sex was also observed, $F(8,\ 1240) = 46.54$, $p < 0.001$, $\eta^2_p = 0.231$. Overall, vocal expressions of happiness ($p < 0.001$) and pleasure ($p < 0.001$) were rated as more pleasant when produced by female than by male speakers, whereas anger ($p < 0.001$), sadness ($p < 0.001$), surprise ($p = 0.032$), and disgust ($p < 0.001$) vocalizations were rated as more unpleasant when produced by female relative to male speakers. Pain vocalizations were rated as more unpleasant when produced by male than by female speakers ($p < 0.001$). No speaker's sex-related differences were observed for neutral and fear vocalizations ($p > 0.050$).

A significant MAV emotional category by listener's sex interaction, $F(8, 1232) = 5.085$, $p < 0.001$, $\eta_p^2 = 0.027$, indicated that female listeners rated anger ($p = 0.004$), sadness ($p < 0.001$), pain ($p = 0.004$), fear ($p = 0.001$) and disgust ($p = 0.018$) vocalizations as more unpleasant compared to male listeners.

Arousal Ratings

A significant main effect of MAV emotional category was observed, $F(8, 1240) = 77.67$, $p < 0.001$, $\eta_p^2 = 0.231$; see Table 5. The highest arousal ratings were observed in the case of vocal expressions of fear ($M = 5.43$), followed by happiness ($M = 5.27$), sadness ($M = 5.21$), anger ($M = 5.12$), pleasure ($M = 5.05$), pain ($M = 5.03$), surprise ($M = 4.73$, disgust ($M = 4.44$), and neutral ($M = 3.30$) vocalizations. Overall, neutral sounds were rated as the least arousing relative to all types of emotional vocalizations ($p < 0.001$ for all comparisons), whereas disgust vocalizations were rated as the least arousing of all emotional vocalizations ($p < 0.001$ for all comparisons, with the exception of surprise—$p = 0.002$). Vocal expressions of happiness were rated as more arousing than surprise ($p = 0.001$) and disgust vocalizations ($p < 0.001$). Vocal expressions of anger were rated as more arousing than surprise ($p < 0.001$), and disgust ($p < 0.001$), and as less arousing than fear ($p < 0.001$). Pleasure vocalizations were rated as more arousing than surprise ($p = 0.020$), and disgust ($p < 0.001$), and as less arousing than fear ($p = 0.007$). Vocal expressions of sadness were rated as more arousing than surprise ($p < 0.001$), and disgust ($p < 0.001$). Surprise vocalizations were rated as more arousing than disgust ($p = 0.002$), and as less arousing than vocal expressions of pain ($p < 0.001$) and fear ($p < 0.001$). Pain vocalizations were rated as more arousing than disgust ($p < 0.001$), and as less arousing than fear ($p < 0.001$). Fear vocalizations were rated as more arousing than disgust ($p < 0.001$).

Furthermore, a significant interaction between MAV emotional category and speaker's sex, $F(8, 1240) = 13{,}29$, $p < 0.001$, $\eta_p^2 = 0.079$, showed that female voices were rated as more arousing than male voices (neutral—$p = 0.003$; pleasure—$p < 0.001$; sadness—$p = 0.009$; surprise—$p = 0.003$; pain—$p < 0.001$; fear—$p < 0.001$; disgust—$p < 0.001$), but not when happiness and anger were communicated ($p > 0.05$).

A significant interaction between MAV emotional category and listener's sex, $F(8, 1232) = 4.397$, $p = 0.001$, $\eta_p^2 = 0.028$, revealed that neutral sounds were rated as more arousing by women than by men ($p = 0.050$).

Dominance Ratings

A main effect of MAV emotional category was observed, $F(8, 1240) = 63.158$, $p < 0.001$, $\eta_p^2 = 0.290$; see Table 6: happy vocalizations were characterized by the highest dominance ratings ($M = 6.45$), followed by pleasure ($M = 5.72$), neutrality ($M = 5.28$), anger ($M = 4.99$), disgust ($M = 4.85$), pain ($M = 4.72$), surprise ($M = 4.65$), fear ($M = 4.52$), and sadness ($M = 4.41$). Pairwise comparisons revealed that dominance ratings were higher for happy vocalizations than for all the other vocal categories ($p < 0.001$ for all comparisons). Neutral vocalizations were rated with lower dominance ratings than pleasure ($p = 0.038$), and with higher dominance ratings than vocal expressions of sadness ($p < 0.001$), surprise ($p < 0.001$), pain ($p = 0.012$), fear ($p = 0.002$), and disgust ($p = 0.008$). Angry vocalizations were characterized by lower dominance ratings than vocal expressions of pleasure ($p < 0.001$), and by higher dominance ratings than sadness ($p < 0.001$), surprise ($p < 0.001$), pain ($p < 0.001$), and fear ($p < 0.001$). Pleasure vocalizations were characterized by higher dominance ratings than sadness ($p < 0.001$), surprise ($p < 0.001$), pain ($p < 0.001$), fear ($p < 0.001$), and disgust ($p < 0.001$). Sadness vocalizations were rated with lower dominance ratings than pain ($p = 0.001$), and disgust ($p < 0.001$). Surprise vocalizations were rated with lower dominance ratings than disgust ($p = 0.029$).

A significant MAV emotional category by speaker's sex, $F(8, 1240) = 11.551$, $p < 0.001$, $\eta_p^2 = 0.069$, revealed higher dominance ratings for pleasure ($p < 0.001$) vocalizations when uttered by a female speaker ($p < 0.001$), and for anger and sadness vocalizations when uttered by a male speaker (anger—$p < 0.010$; sadness—$p < 0.001$).

The Relationship Between Dimensional Ratings and Recognition Accuracy

Standardized β coefficients and t test values are shown in Table 7. Figure 1 shows the association between dimensional ratings and unbiased scores for the regression models that were statistically significant. Recognition accuracy was significantly predicted by valence, arousal, and/or dominance for neutral, $F(3, 152) = 3.135$, $p = 0.027$, $R^2 = 0.058$; happiness, $F(3, 152) = 12.931$, $p < 0.001$, $R^2 = 0.203$; pleasure, $F(3, 152) = 14.629$, $p < 0.001$, $R^2 = 0.224$; sadness, $F(3, 152) = 12.319$, $p < 0.001$, $R^2 = 0.196$; pain, $F(3, 152) = 5.130$, $p = 0.002$, $R^2 = 0.092$; and disgust, $F(3, 152) = 6.631$, $p < 0.001$, $R^2 = 0.116$ categories (see Table 7). However, in the case of neutral vocal expressions,

Table 7 Regressions of dimensional ratings predicting recognition accuracy by emotional category (H_u Scores)

	Valence		Arousal		Dominance	
	β	t	β	t	β	t
Predicting Neutral	0.143	1.809	−0.28	−0.358	0.177	2.213*
Predicting Happiness	0.322	3.967***	−0.57	−0.744	0.233	2.984**
Predicting Pleasure	0.432	4.954***	0.003	0.039	0.100	1.330
Predicting Surprise	0.083	0.997	0.035	0.422	0.005	0.060
Predicting Anger	−0.174	−2.153	0.058	0.720	0.121	1.500
Predicting Sadness	−0.432	−5.588***	0.027	0.339	−0.008	−0.104
Predicting Pain	−0.301	−3.825***	0.022	0.280	0.042	0.529
Predicting Fear	0.024	0.281	0.161	1.883	0.021	0.261
Predicting Disgust	−0.318	−4.147***	0.091	1.186	0.026	0.335

β values, t scores, and significance levels are provided for each emotional dimension with respect to each emotional category; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

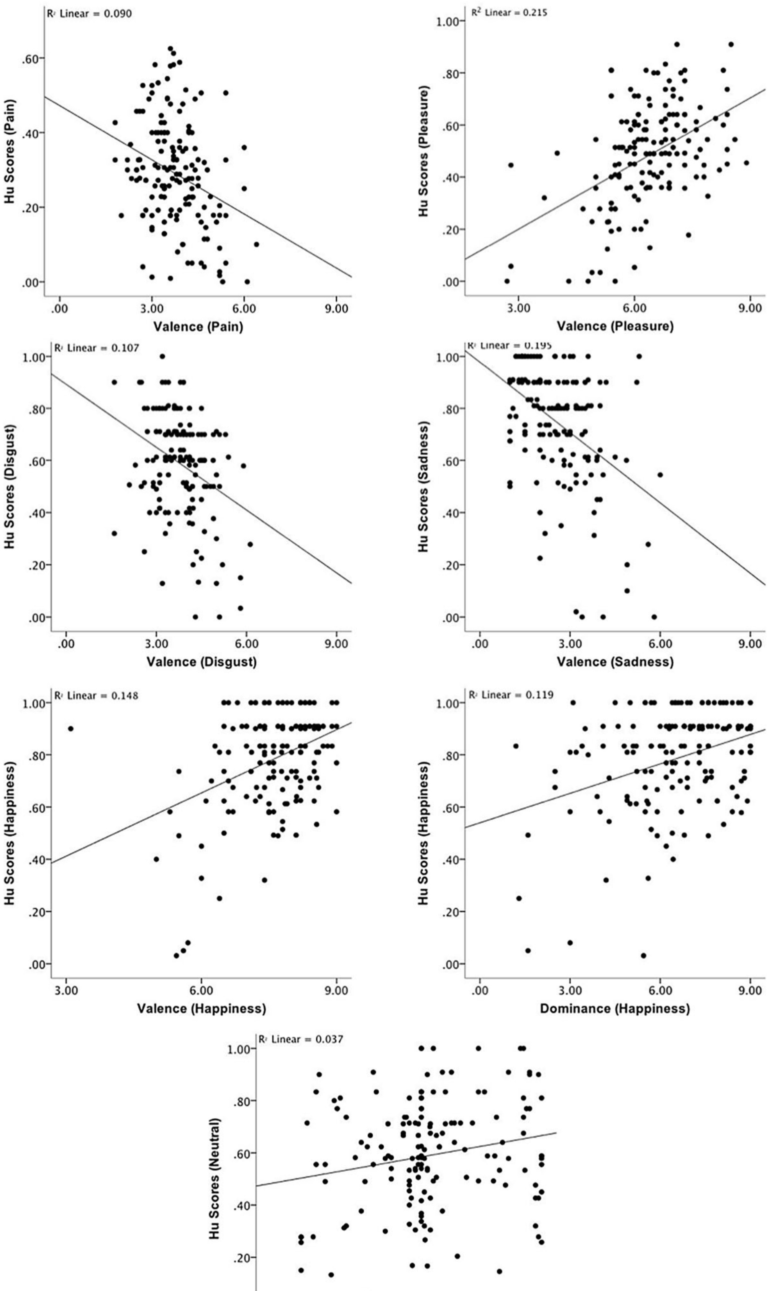

Fig. 1 Graphical representation of the regression analyses (only statistically significant models are presented, $p < 0.05$)

valence and arousal did not add statistically significantly to the prediction ($p > 0.05$); in the case of happiness, arousal did not add statistically significantly to the prediction ($p > 0.05$); in the case of pleasure, sadness, pain, and disgust, arousal and dominance did not add statistically significantly to the prediction ($p > 0.05$). In the case of anger, surprise and fear, none of the affective dimensional ratings provided significant contributions to the prediction of recognition accuracy ($p > 0.05$).

Discussion

This study aimed to clarify how emotions are recognized from nonverbal vocal cues using a forced-choice recognition task. The present investigation is novel in assessing unbiased recognition accuracy for nonverbal emotional vocalizations taken from a widely used stimulus battery such as the MAV. Also, by providing a deeper characterization of the categorical and dimensional structure of the MAV vocal sounds, this study probed which affective dimensions better predict the decoding of discrete emotions.

Above-chance accuracy levels lend support to the idea that adult listeners are adept at extracting emotional information from vocal stimuli, and that nonverbal vocalizations are effective in conveying emotions (Sauter et al. 2010a, b; Schröder 2003). Furthermore, based on the analysis of unbiased accuracy scores (Wagner 1993), the present study corroborates the observation that emotions conveyed by the voice are perceived categorically (e.g., Juslin and Laukka 2003), and that some emotional categories tend to be more easily recognized than others (see, for example, Hawk et al. 2009; Lima et al. 2013).

Unbiased Accuracy in Nonverbal Vocal Emotional Recognition

Whereas surprise was the most difficult emotion to decode from the vocal samples, happiness was the most accurately recognized category in the current study. This finding is in contradiction with the studies of Belin et al. (2008) and Koeda et al. (2013), in which sadness was the most accurately recognized emotion. Further, this finding does not support the claim that happiness is only effectively expressed when communicated through the face (Ekman 1994; Wallbott 1988).

As stated before, the accuracy in decoding emotional meaning may be related to differences in the acoustic distinctiveness of vocal emotions (see Banse and Scherer 1996; Paulmann et al. 2012, for similar results with prosodic stimuli). For example, low acoustic power (dB) values are a distinctive property of happy vocalizations in the MAV battery (Belin et al. 2008), which may have facilitated the decoding of this specific emotion. Besides, it could be argued that happiness recognition benefited from the lower number of positive emotional categories (e.g., happiness and pleasure) compared to the number of negative emotional categories (e.g., anger, pain, disgust, fear, and sadness).

In alternative, the highest recognition accuracy for happy vocal sounds may have been driven by their high social significance. In this respect, it is worth highlighting the fact that the happy vocalizations in the MAV battery consist of laughs. Positive vocal emotions, such as laughter, play a critical role in communication and social bonding (Scott et al. 2014). Specifically, laughter serves a critical social function, by eliciting positive affect in the listeners, modifying its arousal and creating positive learned experiences (Owren and Bachorowski 2003). In line with these findings, Pell et al. (2015) found an early differentiation of happy vocalizations (laughter) in the EEG signal when compared to other vocal emotional categories. Using MAV nonverbal vocalizations as well as pseudo-utterances, the authors reported an early N1 peak and the earliest P2 peak for happy (laughter) vocalizations when compared to all voice/emotion combinations. Furthermore, in close parallel to our findings, in this study (Pell et al. 2015) laughs received high ratings of valence and arousal. It is also possible that the high accuracy in the recognition of happiness observed in the current sample was related to the increased perceived pleasantness of the vocal sounds. Indeed, happy stimuli were, in the current study, associated with the highest valence and dominance ratings, and second highest arousal ratings, which confirms the uniqueness of these vocal signals. It is also worth noting that, even though universally recognized, emotional decoding of positive vocalizations seems to be more sensitive to the effects of culture than negative emotions (e.g., Sauter et al. 2010b). This raises the additional possibility that accuracy in recognizing happy vocal expressions was modulated by specific features of the Portuguese cultural context. However, this hypothesis awaits further investigation.

Contrary to happiness, surprise was the most poorly recognized emotion. This finding contrasts with the study of Belin et al. (2008) in which pain was the most poorly recognized emotion, but it agrees with previous studies that used prosodic pseudospeech and unbiased recognition rates (Liu and Pell 2012). The associated low recognition accuracy may be due to the acoustic similarities between surprise and fear, as well as to its high f0 range (Pell et al. 2009b). In good agreement with the low acoustic distinctiveness between these two vocal emotions (characterized by the shortest durations and the highest median F0), confusion between surprise and fear was also reported in the MAV validation by Belin et al. (2008). Furthermore, a similar confusion pattern was observed in other sensory modalities as well (e.g., faces—Ekman et al. 1972).

The current study also revealed that anger vocalizations were poorly recognized. Considering the evolutionary need of easily attending to and rapidly recognizing angry sounds, a plausible hypothesis was that vocal expressions of anger would be associated with higher accuracy rates. The low accuracy in decoding anger cues may be, once more, due to the acoustic properties of the MAV sounds. There is evidence suggesting acoustic similarities between vocal expressions of anger and of surprise and pain (Belin et al. 2008), a confusion pattern that was corroborated by our data (see supplementary data). The low accuracy rates found for surprise and anger vocalizations in our study [recognized with 75 and 78% accuracy, respectively, in the study of Belin et al. (2008)] recommend caution when using these specific MAV stimuli in emotion research, at least with European Portuguese participants.

Differences in sample sizes (e.g., 29 participants in the study of Belin et al. 2008 vs. 156 participants in the current study) may also account for the discrepant results. Critically, the current study probed vocal emotional recognition using a forced-choice task, whereas Belin et al. (2008) tested recognition accuracy using intensity scores and measured accuracy based on biased hit rates. The apparent discrepancies may be a consequence of these methodological differences. Another possibility is that not all MAV stimuli in the battery were optimally mapped onto discrete emotions. Thereby, weaknesses in the construction of the MAV stimuli (specifically vocal expressions of anger) may account for the differences between the current study and the original findings reported by Belin and collaborators.

We should also note differences between this and previous studies that probed emotional recognition with prosodic speech. For example, in some of these studies, happiness was more poorly recognized compared to other emotion categories (e.g., Banse and

Scherer 1996; Johnstone and Scherer 2000; see Juslin and Laukka 2003 for a review; Liu and Pell 2012), whereas anger was among the most accurately recognized emotions (see Juslin and Laukka 2003 for a review; Liu and Pell 2012). Contrasting emotion decoding from nonverbal vocalizations versus speech prosody, a recent ERP study revealed electrophysiological differences in the processing of anger cues embedded in both types of vocal expressions, demonstrating that the late positive component elicited by anger stimuli was delayed in the case of prosody (Pell et al. 2015). There is also evidence supporting the facilitated recognition of nonverbal vocalizations compared to speech prosody, namely in the case of emotionally negative expressions (Hawk et al. 2009). These differences provide further support for differences in the neurofunctional mechanisms underlying the processing of vocal emotions as a function of stimulus type, i.e., nonverbal vocalizations versus prosodic speech. In other words, the concurrent processing of linguistic and supralinguistic information in the prosodic signal seems to affect how emotional cues are processed and identified (Pell et al. 2015).

Dimensional Properties of Nonverbal Vocal Emotions

Regarding the dimensional assessment of the MAV vocalizations, we observed that distinct vocal emotional categories were characterized by distinct patterns of affective dimensions, such as valence and arousal (e.g., Laukka et al. 2005). Neutral vocal expressions were rated as less pleasant than positive vocalizations, and as more pleasant than negative vocalizations, as well as less arousing than emotional vocalizations. Happiness was associated with the highest valence and dominance ratings. Vocal expressions of fear were rated as the most arousing vocal stimuli, in line with Belin et al. (2008), followed by happiness and sadness. The role of arousal in vocal emotional perception is especially important, as vocal expressions may primarily reflect arousal or activation patterns (e.g., Bachorowski 1999; Pakosz 1983).

The Effects of Sex on Nonverbal Vocal Emotional Processing

Understanding emotions communicated through the voice hinges not only upon who expresses (speaker's sex) but also upon who perceives the emotion (listener's sex). Supporting the study of Belin et al. (2008), nonverbal affective vocalizations were decoded with higher accuracy when produced by female speakers. Further, women and men differed in the recognition of female and male vocalizations, with respect to specific discrete emotions, which supports a robust body of evidence demonstrating sex differences in emotion recognition (examples in emotional vocal processing include: Schirmer et al. 2005a, b; for a review, see Thompson and Voyer 2014; Hall et al. 2016). Specifically, vocal expressions of anger and sadness were more accurately recognized by women, whereas men were better at recognizing surprise. This finding agrees with the observation that women and men perceive vocal emotional information differently (e.g., Besson et al. 2002; Schirmer and Kotz 2003; Schirmer et al. 2002, 2004, 2005b; Thompson and Voyer 2014), and that women outperform men when recognizing emotions from nonverbal cues (Hall 1978, 1984; Hall and Matsumoto 2004; McClure 2000). Specifically, our results provide support to a female advantage in the recognition of negative sounds (e.g., Gohier et al. 2013). For example, previous studies observed higher accuracy rates in women compared to men specifically when labelling angry and sad expressions (reviewed in the meta-analysis by Thompson and Voyer 2014). The female superiority in the recognition of negative vocal emotions supports the "fitness threat hypothesis" (Hampson et al. 2006), according to which the enhanced accuracy shown by females in response to negative emotional stimuli is a result of their evolutionary role as primary caretakers.

Besides the categorical judgments of vocal emotions, dimensional ratings were also modulated by the listener's sex and speaker's sex. Overall, female vocalizations were rated as more arousing than vocalizations produced by male speakers. Nonetheless, this effect was dependent on the emotion category: vocalizations produced by female speakers were rated as more pleasant and with higher dominance in the case of positive vocal expressions only; negative vocalizations (anger and sadness) elicited higher dominance and lower valence ratings when produced by male speakers. Moreover, the effect of listener's sex was significant in the case of valence and arousal, but not dominance. In particular, women rated negative vocalizations (anger, sadness, pain, fear, disgust) as more unpleasant than men, confirming the negativity bias observed in previous studies with Portuguese participants (affective words—Soares et al. 2012; affective sounds—Soares et al. 2013; affective pictures—Soares et al. 2015). These findings lend new support to the role of sex differences in the perception of vocal emotional information (e.g., Besson et al. 2002; Paulmann and Kotz 2008; Schirmer and Kotz 2003; Schirmer et al. 2002, 2004, 2005b).

The Relationship Between Categorical and Dimensional Ratings of Nonverbal Vocal Emotions

Keeping with previous studies (Stevenson and James 2008), regression analyses showed that dimensional ratings do not consistently predict listeners' recognition accuracy. We also observed a lack of homogeneity in the ability of dimensional ratings to predict categorical ratings. For example, dimensional ratings of fear and surprise vocalizations did not predict the recognition accuracy of these discrete emotions. Nonetheless, recognition of happiness, pleasure, sadness, pain, and disgust was consistently predicted by valence ratings. Thus, it seems that valence may properly characterize emotion recognition, but only in the case of specific vocal emotion categories. The lack of consistency may be related to the acoustic properties of the vocalizations, with specific cues being more pertinent for some emotions compared to others, and thereby affecting both dimensional and categorical ratings differently. However, this remains a speculation and needs to be examined in future studies.

Together, these results indicate that dimensional or categorical assessments alone cannot be used to characterize the affective properties of a vocal stimulus (Pinheiro et al., in press; Stevenson et al. 2007).

Strengths and Weaknesses of the Current Study

A major strength of the current study is to provide an unbiased measure of vocal emotional decoding, based on a forced-choice categorization task. Previous studies have used indirect recognition measures such as intensity ratings (Belin et al. 2008; Koeda et al. 2013) or raw hit scores. Our analysis approach, based on H_u scores, takes into account the discrepancies in item frequency among the categories and the biases in listeners' responses, such as the preference for specific response alternatives. In addition to that, the current study collected data from a large sample of subjects ($N = 156$) as opposed to previous studies that aimed to probe emotion decoding from nonverbal vocalizations (Belin et al. 2008; Hawk et al. 2009; Koeda et al. 2013; Lima et al. 2013). Nonetheless, these findings are specific of the stimuli set used (MAV), and therefore their generalization to other vocal samples might be limited.

Conclusions

By allowing a more complete characterization of the categorical and dimensional structure of the MAV sounds, the present study offers researchers a new avenue to develop well-controlled studies intended to probe vocal emotional processing. Our data indicate happiness as the emotional category that is more easily recognized when emotions are expressed through nonverbal vocalizations. This finding supports the notion that laughter plays a highly relevant social function in nonverbal communication. The present study also supports the need for combining categorical and dimensional accounts when characterizing emotional vocal sounds. Additionally, the current findings indicate that both speaker's and listener's sex affect how accurately vocal emotions are judged and recognized.

By understanding how humans perceive and recognize the emotional 'melody' of the voice, we will become closer to a deeper understanding of the dynamics of social communication.

Acknowledgments The authors gratefully acknowledge all the participants who collaborated in the study.

Funding This work was supported by Grants IF/00334/2012, and PTDC/MHN-PCN/3606/2012 funded by *Fundação para a Ciência e a Tecnologia* (FCT, Portugal) and *Fundo Europeu de Desenvolvimento Regional* (FEDER) through the European programs *Quadro de Referência Estratégico Nacional* (QREN), and *Programa Operacional Factores de Competitividade* (COMPETE), awarded to A.P.P., and by a Doctoral Grant (SFRH/BD/52400/2013) funded by *Fundação para a Ciência e a Tecnologia* (Portugal), awarded to M.V.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

References

Bachorowski, J. A. (1999). Vocal expression and perception of emotion. *Current Directions in Psychological Science, 8,* 53–57. doi:10.1111/1467-8721.00013.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70,* 614–636. doi:10.1037/0022-3514.70.3.614.

Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences, 8,* 129–135. doi:10.1016/j.tics.2004.01.008.

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,* 531–539. doi:10.3758/BRM.40.2.531.

Besson, M., Magne, C., & Schön, D. (2002). Emotional prosody: Sex differences in sensitivity to speech melody. *Trends in Cognitive Sciences, 6,* 405–407. doi:10.1016/S1364-6613(02)01975-7.

Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001a). Emotion and motivation I: Defensive and appetitive reactions in picture processing. *Emotion, 1*(3), 276. doi:10.1037/1528-3542.1.3.276.

Bradley, M. M., Codispoti, M., Sabatinelli, D., & Lang, P. J. (2001b). Emotion and motivation II: Sex differences in picture processing. *Emotion, 1,* 300–319. doi:10.1037/1528-3542.1.3.300.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25,* 49–59. doi:10.1016/0005-7916(94)90063-9.

Bradley, M. M., & Lang, P. J. (1999). *The International Affective Digitized Sounds (IADS): Stimuli. Instruction manual and affective ratings*. Florida. FL: The Center for Research in Psychophysiology. University of Florida.

Bradley, M. M., & Lang, P. J. (2007). *The International Affective Digitized Sounds (IADS-2): Affective ratings of sounds and instruction manual* (2nd ed.). Gainesville: NIMH Center for the Study of Emotion and Attention, University of Florida.

Darwin, C. (1998). *The expression of the emotions in man and animals* (3rd ed.). London: Harper-Collins. **(Original work published in 1872)**.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6,* 169–200. doi:10.1080/02699939208411068.

Ekman, P. (1994). Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique. *Psychological Bulletin, 115,* 268–287.

Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). *Emotion in the human face: Guide-lines for research and an integration of findings*. New York, NY: Pergamon.

Fecteau, S., Belin, P., Joanette, Y., & Armony, J. L. (2007). Amygdala responses to nonlinguistic emotional vocalizations. *Neuroimage, 36*(2), 480–487. doi:10.1016/j.neuroimage.2007.02.043.

Gohier, B., Senior, C., Brittain, P. J., Lounes, N., El-Hage, W., Law, V., et al. (2013). Gender differences in the sensitivity to negative stimuli: Cross-modal affective priming study. *European Psychiatry, 28,* 74–80. doi:10.1016/j.eurpsy.2011.06.007.

Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin, 85,* 845–857. doi:10.1037/0033-2909.85.4.845.

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

Hall, J. A., Andrzejewski, S. A., Murphy, N. A., Mast, M. S., & Feinstein, B. A. (2008). Accuracy of judging others' traits and states: Comparing mean levels across tests. *Journal of Research in Personality, 42*(6), 1476–1489. doi:10.1016/j.jrp.2008.06.013.

Hall, J. A., Gunnery, S. D., & Horgan, T. G. (2016). Gender differences in interpersonal accuracy. In J. A. Hall, M. Schmid Mast, & T. V. West (Eds.), *The social psychology of perceiving others accurately* (pp. 309–327). Cambridge: Cambridge University Press.

Hall, J. A., & Matsumoto, D. (2004). Gender differences in judgments of multiple emotions from facial expressions. *Emotion, 4*(2), 201. doi:10.1037/1528-3542.4.2.201.

Hampson, E., van Anders, S. M., & Mullin, L. I. (2006). A female advantage in the recognition of emotional facial expressions: Test of an evolutionary hypothesis. *Evolution and Human Behavior, 27,* 401–416. doi:10.1016/j.evolhumbehav.2006.05.002.

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. doi:10.1037/a0015178.

Jiang, X., Paulmann, S., Robin, J., & Pell, M. D. (2015). More than accuracy: Nonverbal dialects modulate the time course of vocal emotion recognition across cultures. *Journal of Experimental Psychology: Human Perception and Performance, 41*(3), 597. doi:10.1037/xhp0000043.

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

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129,* 770–814. doi:10.1037/0033-2909.129.5.770.

Kensinger, E. A., & Corkin, S. (2004). Two routes to emotional memory: Distinct neural processes for valence and arousal. *Proceedings of the National Academy of Sciences of the United States of America, 101*(9), 3310–3315. doi:10.1073/pnas.0306408101.

Koeda, M., Belin, P., Hama, T., Masuda, T., Matsuura, M., & Okubo, Y. (2013). Cross-cultural differences in the processing of non-verbal affective vocalizations by Japanese and Canadian listeners. *Frontiers in Psychology, 4,* 105. doi:10.3389/fpsyg.2013.00105.

Kotchoubey, B., Kaiser, J., Bostanov, V., Lutzenberger, W., & Birbaumer, N. (2009). Recognition of affective prosody in brain-damaged patients and healthy controls: A neurophysiological study using EEG and whole-head MEG. *Cognitive Affective and Behavioral Neuroscience, 9,* 153–167. doi:10.3758/CABN.9.2.153.

Kotz, S. A., Kalberlah, C., Bahlmann, J., Friederici, A. D., & Haynes, J. D. (2013). Predicting vocal emotion expressions from the human brain. *Human Brain Mapping, 34*(8), 1971–1981. doi:10.1002/hbm.22041.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International Affective Picture System (IAPS): Technical manual and affective ratings*. Gainesville. FL: Center for Research in Psychophysiology. University of Florida.

Laukka, P. (2005). Categorical perception of vocal emotion expressions. *Emotion, 5,* 277–295. doi:10.1037/1528-3542.5.3.277.

Laukka, P., Elfenbein, H. A., Söder, N., Nordström, H., Althoff, J., Chui, W., et al. (2013). Cross-cultural decoding of positive and negative non-linguistic emotion vocalizations. *Frontiers in Psychology, 4,* 353. doi:10.3389/fpsyg.2013.00353.

Laukka, P., Juslin, P., & Bresin, R. (2005). A dimensional approach to vocal expression of emotion. *Cognition and Emotion, 19,* 633–653. doi:10.1080/02699930441000445.

Lewis, P. A., Critchley, H. D., Rotshtein, P., & Dolan, R. J. (2007). Neural correlates of processing valence and arousal in affective words. *Cerebral Cortex, 17*(3), 742–748. doi:10.1093/cercor/bhk024.

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. doi:10.1037/a0034287.

Lima, C. F., Castro, S. L., & Scott, S. K. (2013). When voices get emotional: A corpus of nonverbal vocalizations for research on emotion processing. *Behavior Research Methods, 45,* 1234–1245. doi:10.3758/s13428-013-0324-3.

Liu, P., & Pell, M. D. (2012). Recognizing vocal emotions in Mandarin Chinese: A validated database of Chinese vocal emotional stimuli. *Behavior Research Methods, 44*(4), 1042–1051. doi:10.3758/s13428-012-0203-3.

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,* 108–112. doi:10.1097/WNR.0b013e32834ea757.

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,* 424–453. doi:10.1037/0033-2909.126.3.424.

Morris, J. S., Scott, S. K., & Dolan, R. J. (1999). Saying it with feeling: Neural responses to emotional vocalizations. *Neuropsychologia, 37*(10), 1155–1163. doi:10.1016/S0028-3932(99)00015-9.

Owren, M. J., & Bachorowski, J. A. (2003). Reconsidering the evolution of nonlinguistic communication: The case of laughter. *Journal of Nonverbal Behavior, 27*(3), 183–200. doi:10.1023/A:1025394015198.

Pakosz, M. (1983). Attitudinal judgments in intonation: Some evidence for a theory. *Journal of Psycholinguistic Research, 12,* 311–326. doi:10.1007/BF01067673.

Paulmann, S., Jessen, S., & Kotz, S. A. (2012). It's special the way you say it: An ERP investigation on the temporal dynamics of two types of prosody. *Neuropsychologia, 50,* 1609–1620. doi:10.1016/j.neuropsychologia.2012.03.014.

Paulmann, S., & Kotz, S. A. (2008). Early emotional prosody perception based on different speaker voices. *NeuroReport, 19,* 209–213. doi:10.1097/WNR.0b013e3282f454db.

Paulmann, S., & Pell, M. D. (2011). Is there an advantage for recognizing multi-modal emotional stimuli? *Motivation and Emotion, 35*(2), 192–201. doi:10.1007/s11031-011-9206-0.

Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009a). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior, 33*(2), 107–120. doi:10.1007/s10919-008-0065-7.

Pell, M. D., Paulmann, S., Dara, C., Alasseri, A., & Kotz, S. A. (2009b). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics, 37,* 417–435. doi:10.1016/j.wocn.2009.07.005.

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. doi:10.1016/j.biopsycho.2015.08.008.

Pinheiro, A. P., Barros, C., & Pedrosa, J. (2016). Salience in a social landscape: Electrophysiological effects of task-irrelevant and infrequent vocal change. *Social Cognitive and Affective Neuroscience, 11*(1), 127–139. doi:10.1093/scan/nsv103.

Pinheiro, A. P., Del Re, E., Mezin, J., Nestor, P. G., Rauber, A., McCarley, R. W., et al. (2013). Sensory-based and higher-order operations contribute to abnormal emotional prosody processing in schizophrenia: An electrophysiological investigation. *Psychological Medicine, 43,* 603–618. doi:10.1017/S003329171200133X.

Pinheiro, A. P., Dias, M., Pedrosa, J., & Soares, A. P. (in press). Minho Affective Sentences (MAS): Probing the role of sex, mood and empathy in affective ratings of verbal stimuli. *Behavior Research Methods*. doi:10.3758/s13428-016-0726-0.

Pinheiro, A. P., Rezaii, N., Rauber, A., Liu, T., Nestor, P. G., McCarley, R. W., et al. (2014). Abnormalities in the processing of emotional prosody from single words in schizophrenia. *Schizophrenia Research, 152,* 235–241. doi:10.1016/j.schres.2013.10.042.

Rosenthal, R., & Rubin, D. B. (1989). Effect size estimation for one-sample multiple-choice-type data: Design, analysis, and meta-analysis. *Psychological Bulletin, 106*(2), 332. doi:10.1037/0033-2909.106.2.332.

Sauter, D. A., & Eimer, M. (2010). Rapid detection of emotion from human vocalizations. *Journal of Cognitive Neuroscience, 22,* 474–481. doi:10.1162/jocn.2009.21215.

Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010a). Perceptual cues in nonverbal vocal expressions of emotion. *The Quarterly Journal of Experimental Psychology, 63,* 2251–2272. doi:10.1080/17470211003721642.

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010b). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences, 107*(6), 2408–2412. doi:10.1073/pnas.0908239106.

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. doi:10.1111/j.2044-835X.2012.02081.x.

Sauter, D. A., & Scott, S. K. (2007). More than one kind of happiness: Can we recognize vocal expressions of different positive states? *Motivation and Emotion, 31*(3), 192–199. doi:10.1007/s11031-007-9065-x.

Scherer, K. R., & Ellgring, H. (2007). Multimodal expression of emotion: Affect programs or componential appraisal patterns? *Emotion, 7,* 158–171. doi:10.1037/1528-3542.7.1.113.

Scherer, K. R., Ladd, D. R., & Silverman, K. E. (1984). Vocal cues to speaker affect: Testing two models. *The Journal of the Acoustical Society of America, 76,* 1346–1356. doi:10.1121/1.391450.

Schirmer, A., & Kotz, S. A. (2003). ERP evidence for a sex-specific Stroop effect in emotional speech. *Journal of Cognitive Neuroscience, 15,* 1135–1148. doi:10.1162/089892903322598102.

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2002). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research, 14,* 228–233. doi:10.1016/S0926-6410(02)00108-8.

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2005a). On the role of attention for the processing of emotions in speech: Sex differences revisited. *Cognitive Brain Research, 24*(3), 442–452. doi:10.1016/j.cogbrainres.2005.02.022.

Schirmer, A., Striano, T., & Friederici, A. D. (2005b). Sex differences in the preattentive processing of vocal emotional expressions. *NeuroReport, 16,* 635–639. doi:10.1097/00001756-200504250-00024.

Schirmer, A., Zysset, S., Kotz, S. A., & von Cramon, D. Y. (2004). Gender differences in the activation of inferior frontal cortex during emotional speech perception. *NeuroImage, 21,* 1114–1123. doi:10.1016/j.neuroimage.2003.10.048.

Schröder, M. (2003). Experimental study of affect bursts. *Speech Communication, 40,* 99–116. doi:10.1016/S0167-6393(02)00078-X.

Scott, S. K., Lavan, N., Chen, S., & McGettigan, C. (2014). The social life of laughter. *Trends in Cognitive Sciences, 18*(12), 618–620. doi:10.1016/j.tics.2014.09.002.

Soares, A. P., Comesaña, M., Pinheiro, A. P., Simões, A., & Frade, C. S. (2012). The adaptation of the Affective Norms for English words (ANEW) for European Portuguese. *Behavior Research Methods, 44,* 256–269. doi:10.3758/s13428-011-0131-7.

Soares, A. P., Pinheiro, A. P., Costa, A., Frade, C. S., Comesaña, M., & Pureza, R. (2013). Affective auditory stimuli: Adaptation of the international affective digitized sounds (IADS-2) for European Portuguese. *Behavior Research Methods, 45,* 1168–1181. doi:10.3758/s13428-012-0310-1.

Soares, A. P., Pinheiro, A. P., Costa, A., Frade, C. S., Comesaña, M., & Pureza, R. (2015). Adaptation of the International Affective Picture System (IAPS) for European Portuguese. *Behavior Research Methods, 47*(4), 1159–1177. doi:10.3758/s13428-014-0535-2.

Stevenson, R. A., & James, T. W. (2008). Affective auditory stimuli: Characterization of the International Affective Digitized Sounds (IADS) by discrete emotional categories. *Behavior Research Methods, 40,* 315–321. doi:10.3758/BRM.40.1.315.

Stevenson, R. A., Mikels, J. A., & James, T. W. (2007). Characterization of the affective norms for English words by discrete emotional categories. *Behavior Research Methods, 39,* 1020–1024. doi:10.3758/BF03192999.

Thompson, A. E., & Voyer, D. (2014). Sex differences in the ability to recognize non-verbal displays of emotion: A meta-analysis. *Cognition and Emotion, 28,* 1164–1195. doi:10.1080/02699931.2013.875889.

Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior, 17*(1), 3–28. doi:10.1007/BF00987006.

Wallbott, H. G. (1988). Big girls don't frown, big boys don't cry—Gender differences of professional actors in communicating emotion via facial expression. *Journal of Nonverbal Behavior, 12,* 98–106.