Universidade do Minho
Escola de Psicologia

Augusto José Martins Mendes

# Recognizing non-verbal emotional vocalizations through the life span: a validated sound database

outubro de 2016

Universidade do Minho
Escola de Psicologia

Augusto José Martins Mendes

# Recognizing non-verbal emotional vocalizations through the life span: a validated sound database

Dissertação de Mestrado

Mestrado Integrado em Psicologia

Trabalho efetuado sob a orientação da

**Professora Doutora Ana P. Teixeira Pinheiro**

outubro de 2016

DECLARAÇÃO

**Nome:** Augusto José Martins Mendes

**Endereço Eletrónico:** a65563@alunos.uminho.pt

**Número do cartão do cidadão:** 14347488

**Título da Dissertação:** Recognizing non-verbal emotional vocalizations through the life span: a validated sound database

**Orientador:** Professora Doutora Ana P. Teixeira Pinheiro

**Ano de conclusão:** 2016

**Designação do Mestrado:** Mestrado Integrado em Psicologia



Universidade do Minho, ____/____/____

Assinatura:_____

**INDEX**

## Agradecimentos

À Professora Ana Pinheiro por toda a disponibilidade e atenção durante o projeto, onde a sua ajuda foi fundamental para a finalização e aperfeiçoamento da Dissertação. A exigência e o cuidado foram essenciais para o meu crescimento como estudante e futuro profissional, encontrando um exemplo a seguir para o meu futuro.

Aos meus colegas da equipa de investigação pela interajuda, principalmente o Diogo por todo o companheirismo e amizade. Todo este trajeto tornou-se mais fácil quando é percorrido em conjunto.

A todos os meus colegas de curso, que percorreram esta etapa comigo desde o primeiro momento que pisei esta Universidade. Principalmente à Dani, Moita, Marco, Tânia e Filipe. Todos os momentos partilhados desde o primeiro dia serão relembrados, sempre!

À minha mãe por estar sempre comigo, por transmitir-me uma mensagem de trabalho e paz, sem ti não era possível.

Ao meu pai por todo o esforço que deposita em mim, só me resta provar que toda a tua força não é, nem vai ser, em vão. Sem ti nada disto era possível.

Aos meus amigos, a minha segunda família. Vocês não precisam dos vossos nomes aqui, vocês sabem quem são. Também não precisam de palavras de agradecimento, sabem bem a importância que têm para mim.

À Inês, por estar sempre do meu lado e conseguir aturar-me sempre com um carinho especial, nos bons e maus momentos. A tua ajuda foi fundamental durante a realização da Dissertação, principalmente quando tudo parecia mal, tu fizeste parecer que tudo estava bem. E por isso, vou estar sempre agradecido.

*"Sabemos muito mais do que julgamos, podemos muito mais do que imaginamos."*

José Saramago

Recognizing non-verbal emotional vocalizations through the life span: a validated sound database

Abstract

Non-verbal emotional vocalizations are sounds (e.g., laughter or cry), which although do not have verbal content, can transmit emotions. These are present since the beginning of life-span (e.g., babies cry), as well in the commencement of human being, even before language appearance. This study pretends to study vocalizations, exploring different age groups (i.e., children, adolescents, young adults, adults and elderly), divided equally for both sexes. Emotional categories in study are: achievement, amusement, pleasure, relief, surprise, anger, disgust, fear and neutral. Initially, in order to collect stimuli, was performed recording sessions with the studied age groups, totaling 200 vocalizations selected. Posteriorly, in order to study the different conditions adjacent to the stimuli, vocalizations were presented to young adults ($n$ = 48) and children ($n$ = 36), who had to categorize each vocalization in a forced choice task, as well as rate according to a valence scale. Main results showed a higher recognition accuracy in vocalizations conveyed by children, as well as female speakers. Regarding emotional categories, disgust and amusement were better recognized, on the other hand fear and achievement had the lowest accuracy rates.

*Key-words:* non-verbal emotional vocalizations; sex; age; sound database.

Reconhecimento de vocalizações emocionais não verbais ao longo do ciclo de vida: validação de uma bateria de sons

Resumo

Vocalizações emocionais não-verbais são sons (e.g., riso ou choro), que se apesar de não conterem conteúdo verbal, transmitem emoções. Estas estão presentes desde o início da nossa vida (e.g., o choro nos bebés), assim como nos primórdios do ser humano, quando a linguagem ainda não tinha surgido. O presente estudo pretende estudar vocalizações, explorando diferentes categorias emocionais e locutores de diferentes faixas etárias (i.e., crianças, adolescentes, jovens adultos, adultos e idosos), divididos equitativamente pelos dois sexos. As categorias emocionais em estudo são: realização/triunfo, diversão, prazer, alívio, surpresa, raiva, nojo, medo, tristeza e neutro. Primeiramente, de modo a obter os estímulos, foram realizadas gravações com as faixas etárias em questão, resultando num total de 200 vocalizações selecionadas. Posteriormente, de modo a estudar as diferentes condições adjacentes ao estímulo, as vocalizações foram apresentadas a jovens adultos ($n$ = 48) e crianças ($n$ = 36), que tiveram que categorizar cada vocalização numa tarefa *forced choice*, assim como, avaliar numa escala de valência. Principais resultados remetem para um maior reconhecimento emocional em vocalizações proferidas por crianças, assim como em locutoras femininas. As categorias emocionais mais reconhecidas foram o nojo e diversão, enquanto que medo e realização obtiveram a pior taxa de eficácia de reconhecimento.

*Palavras-chave:* vocalizações emocionais não verbais; sexo; idade; bateria de sons.

## Introduction

Imagine that you are speaking on the phone with someone you have never met before or heard his or her voice. Although it is the first time you hear that voice, you can immediately identify numerous characteristics of the speaker, such as his/her sex, age, emotional state, among others (Amilon, van de Weijer, & Schötz, 2007; Belin, Fecteau, & Bédard, 2004). The ability to decode several characteristics of the speaker relies on the configuration of different acoustic cues (e.g., pitch contour, intensity, fundamental frequency - f0), which determine how we decode emotions and how we communicate with others (Juslin & Laukka, 2003; Murray & Arnott, 1993; Sauter, Eisner, Calder, & Scott, 2010).

Our communicative systems do not depend only on lexical codes (that are culturally dependent) but also on non-verbal information such as vocal emotional signals, which tend to be cross-culturally recognized with similar accuracy, representing examples of psychological universals (Sauter, Eisner, Ekman, & Scott, 2010). The existing research has focused more on the study of emotional facial expressions (e.g., Ekman, 1992), and fewer studies have investigated vocal emotions (e.g., Juslin & Laukka, 2003).

Emotions have been explained by different approaches due to its complexity, independently of its communication channel (e.g., facial expressions, voice). Two main perspectives should be highlighted: the categorical and dimensional approaches (Fujimura, Matsuda, Katahira, Okada, & Okanoya, 2016). On the one hand, the dimensional approach suggests that emotional cues should be conceptualized along different affective dimensions, such as arousal and valence. This model is supported by recent findings from behavioral, neurocognitive and developmental studies (Posner, Russell, & Peterson, 2005). Valence is defined as the level of subjective experience that ranges from pleasantness to unpleasantness (Laukka, Juslin, & Bresin, 2005; Russell, 1980). Emotional cues may also be classified according to their in arousal, which represents how much activation is felt, which ranges from not aroused to extremely aroused (Russell, 1980). On the other hand, the categorical approach advocates the existence of a limited number of basic emotions (e.g., anger, disgust, fear, happiness, sadness and surprise), which are considered universal (Ekman, 1992; Sauter, Eisner, Ekman, et al., 2010).

**The role of non-verbal emotional vocalizations in communication**

In the realm of the voice, it is possible to distinguish two main channels of emotional expression: the "affective bursts", also known as non-verbal emotional vocalizations (Belin, Filllion-Bilodeau, & Gosselin, 2008; Lima, Castro, & Scott, 2013; Sauter, 2007; Schroder, 2003), and emotional speech prosody (Paulmann, Pell, & Kotz, 2008; Wiethoff, Wildgruber, & Becker, 2008).

The importance of non-verbal vocalizations in emotional processing is demonstrated in several domains, including an evolutionary approach. Ontogenetically, non-verbal communication through vocalizations preceded language, which were evolved across different brain areas, in order to suit distinct functions, allowing a great channel used to survive and communicate before language appearance, as well as phylogenetically (Belin et al., 2004; Owren, Amoss, & Rendall, 2011). For instance, in babies, while language is not yet developed (e.g. laughs, cry or screams), vocalizations also allow the expression of a few needs, which communicate crucial information about the affective state of the infant (Belin et al., 2004). Moreover, infants produce spontaneous non-verbal vocalizations, such as crying or laughter, even if they are deaf or blind, suggesting that this type of communication is innate (Owren, Amoss, & Rendall, 2011). Furthermore, speech emerged when our brain was already able to process different types of vocal information, which suggests that the processing of affective and speech information conveyed by the voice are processed by different partially dissociated functional pathways (Belin et al., 2004; Juslin & Laukka, 2003).

Despite the fact that verbal expressions of emotion have a crucial role in our social interactions, it is known that nonverbal emotional cues are also of vast significance. Scherer introduced the concept of "*affect burst"*, which can be defined as "very brief, discrete, nonverbal expressions of affect in both face and voice as triggered by clearly identifiable events" (cited in Schröder, 2003; Scherer, 1994, p.170). In the same definition, it is clarified that verbal interjections (e.g. "Heaven!") are not considered as affect bursts due to their verbal content. At the same time, affect bursts denote a range of emotional expressions, such as 'raw affect bursts' that are reflexive and universal (e.g., laughter), and 'affect emblems' that are conventionalized symbols and culturally dependent, such as "Yuck!" for disgust (Schröder, 2003). Based on this concept, Sauter (2007) only focused on the processing of nonverbal expressions of emotion containing minimal phonemic information, and labeled them as "non-verbal vocalizations of emotions" meeting this study purpose.

Current literature about vocal emotional processing has been focusing mainly in speech prosody, whereas studies are lacking on non-verbal vocalizations in comparison with others channels (e.g., speech prosody, facial expressions). However, vocalizations have been shown as a robust channel to communicate emotions (Hawk, Kleef, Fischer, & Schalk, 2009), which is not dependent of vision (Schröder, 2003), and it was the first communication mean in human being, which is similar among others primates (Owren et al., 2011).

## How accurate are we in recognizing vocal emotions?

Non-verbal emotional vocalizations are an effective mean to express emotions, as well as facial expressions and prosody, although vocalizations had slightly better recognition accuracy in comparison with the other channels (Hawk et al., 2009). Moreover, non-verbal vocalizations are known to be more easily processed compared to speech prosody (Pell et al., 2015).

Vocalizations, as evolutionary primitive signals, are an excellent channel to express emotions. To explore this idea, Schröder (2003) studied recognition accuracy across 10 different emotions through affect bursts: admiration, threat, disgust, elation, boredom, relief, startle, worry, contempt, and hot anger. An overall recognition rate of 81% was obtained exclusively with audio and without context, suggesting that emotions are accurately recognized when conveyed through nonverbal vocalizations. In line with this work, a study conducted by Sauter (2007) with other 10 emotions (achievement/triumph, amusement, anger, contentment, disgust, fear, sensual pleasure, relief, sadness, and surprise) demonstrated that emotions were recognized highly above chance (mean rate of 71%).

Moreover, vocalizations are also recognized above chance among children with different ages (i.e., from 5 to 10 years old), with an overall mean of 83.9%, in the following emotional categories: amusement, contentment, relief, achievement, anger, disgust, fear, sadness, surprise and neutral (Sauter, Panattoni, & Happ, 2013).

Vocal emotional studies can be restricted by scarcity of controlled and validated stimuli, and because of this validated sets of non-verbal vocalizations were done. The Montreal Affective Voices (MAV) is a database that includes five negative emotions (anger, disgust, fear, pain and sadness) and two positive emotions (happiness and pleasure), besides surprise and neutral expressions, using the French vowel “ah” (Belin et al., 2008). These emotional categories were associated with a mean recognition rate of 68%.

A more recent validated database tried to resemble more closely real-life vocalizations, allowing all linguistic sounds, but still without any verbal content (e.g., interjections, words), and not only the vowel "ah" like Belin et al. (2008) (Lima et al., 2013). The battery includes four negative (anger, fear, disgust and sadness) and four positive (achievement, amusement, pleasure and sensual pleasure) emotions (Lima et al., 2013). These vocalizations were highly recognized according to the intended emotion, with an overall accuracy of 86%.

Regarding the emotions available in these studies, both presented a higher recognition rate for disgust than the other emotions, which suggests that this emotion is easily recognizable from non-verbal vocalizations (Belin et al., 2008; Lima et al., 2013; Sauter, Eisner, Calder, et al., 2010). This contradicts the results observed in the context of emotional prosody, in which disgust presented the lowest hit rate among the various emotions (Castro & Lima, 2010). These incongruent results might be evidence for a segregation in the neural mechanisms involved in emotion recognition in the auditory modality, i.e. for stimuli with and without verbal content (Belin et al., 2004). On the other hand, fearful vocalizations had the lowest hit rates (Lima et al., 2013). This can be explained by the confusion between fear and surprise, which share acoustic similarities (Belin et al., 2008), but also for its intrinsic meaning (i.e. some fear expressions could correspond at the same time to surprise, and vice-versa, for instance if suddenly you see a snake, it will elicit a fear response, but at the same time it was an unexpected situation, which also can elicit a surprise response), because this confusion is also observed even in facial expression (cited in Belin et al., 2008; Ekman, Friesen, & Ellsworth, 1972).

**The effects of sex on emotional recognition**

There is consistent evidence suggesting that both speaker's sex and listener's sex affect how vocal emotions are decoded. This was explained initially by the primary caretaker hypothesis stating that women's advantage in recognizing emotions is due to their responsibility in child rearing (Babchuk, Hames, & Thompson, 1985). This difference forced adaptations in order to enhance the probability of survival of their offspring and attend to their needs. Nevertheless, this theory has some criticisms and it has been suggested that childcare experience is not the only explanation for this advantage, but also some mechanisms involved in threat situations, however this was specific to facial processing (Hampson, Van Anders, &

Mullin, 2006). Additionally, regarding emotional stimulus processing, women tend to present greater brain and physiological activation in response to negative stimuli than men, while men have greater activation for positive stimuli, particularly erotic stimuli (Stevens, Hamann, Stevens, & Hamann, 2012).

This was also observed in non-verbal vocalizations, because female participants were better decoding emotions than males (Belin et al., 2008). At the same time, speaker's sex also modulates the accuracy of vocal emotional recognition. When non-verbal vocalizations are uttered by female speakers they are better recognized than the ones uttered by male speakers (Belin et al., 2008).

**The effects of age on emotional recognition**

Emotional processing, namely emotional recognition, is in constant change throughout the lifespan. There are some theories that tried to account for age effects in the capacity to decode emotions from visual and auditory stimuli (Mill, Allik, Realo, & Valk, 2009; Phillips, MacLean, & Allen, 2002). Sociocognitive theories propose that an increased ability to process emotions is acquired with age due to the wide life experience resulting from interpersonal communication (Phillips et al., 2002). The cognitive aging approach ensures that cognitive operations, such as discrimination of acoustic details or comparisons with prototypes stored in memory, are needed for regular emotion recognition (Mill et al., 2009). On the other hand, the neuropsychological theory predicts an age-related decline in understanding emotions starting in adulthood, which is explained by the fact that particular regions of the brain involved in the processing of emotion are the same as those affected by normal aging (Phillips et al., 2002; Ruffman, Henry, Livingstone, & Phillips, 2008).

However, the previous theories do not include evidence about younger ages (i.e., children and adolescents) (Mill et al., 2009). Although children are able to decode emotions through vocalizations as referred above (Sauter et al., 2013), they have a worse performance in vocal emotional tasks in comparison with adolescents, with ages above 11 years old (Tonks, Williams, Frampton, Yates, & Slater, 2007). These differences can occur, because in adolescence there is an increase of social interactions, and a consequent improvement of social skills (Turkstra, 2000).

**The current study and hypotheses**

The aim of this study was to study how different characteristic of the speaker can influence emotional recognition and valence assessment. Additionally, a validation of a database of nonverbal vocalizations, due to the scarcity of controlled and validated stimuli in current literature. Other problem observed is the absence of validated stimuli conveyed by different ages.

The emotional recognition of 10 emotional modalities was tested (i.e., sadness, fear, disgust, anger, surprise, relief, pleasure, amusement, achievement, and neutral). These emotional categories were uttered by five distinct age groups (i.e., children, adolescents, young adults, adults, and elderly), divided equally by speakers' sex. To the best of our knowledge, this is the first study to test the effects of speaker's age on vocal emotional recognition, as well speaker's sex within each age group, providing controlled stimuli uttered by distinct age groups.

Regarding participant's age, literature is also scarce about age influence in non-verbal vocalizations recognition, especially among children (Sauter et al., 2013). So, this enhanced the need to study different age groups in the validation task in the present study: children and young adults. In order to achieve this goal, the first stage of the study was to record the voice of subjects with different ages. Then, the vocal samples illustrating the 10 emotional categories were presented to children and young adult participants. Participants had to assess each emotion, according to categorical and valence assessment.

We expected young adults and children to demonstrate an overall high recognition rate, higher than chance (Belin et al., 2008; Lima et al., 2013; Sauter et al., 2013; Schroder, 2003). However, we hypothesized young adults would show better performance in recognition due to the higher maturity of the brain and cognitive processes, which happens during adolescence (Phillips et al., 2002; Tonks et al., 2007; Turkstra, 2000). We also expected female listeners to have a higher accuracy in comparison with males (Belin et al., 2008; Schirmer, Kotz, & Friederici, 2002)

Differences in vocal emotional recognition were also expected in both groups as a function of speakers' age and sex. As in previous studies, it's expected vocalizations conveyed by female speakers as better recognized (Belin et al., 2008). Moreover, we expected recognition accuracy to be higher for vocalizations produced by speakers of the same age group of the participants, even though there is no existing evidence to support this hypothesis.

Non-verbal vocalizations are a recent subject in this area, despite being one of the ancient processes in human species. Regarding emotional processing, literature had focused mostly in facial expressions, including cross-sectional studies, pointing this study as pertinent and innovative, which can raise new paradigms and perspectives.

## Method

The study involved two distinct stages with different participant samples: 1) voice recording, and 2) ratings of vocalizations, which are described below. This study was approved by the Ethics Commission of the University of Minho.

### Stage 1: Voice Recording

**Participants.** Thirty-five European Portuguese participants were recruited for each age group (children: 8-10 years, adolescents: 14-16 years, young adults: 19-23 years, adults: 45-52 years, and older adults: >65 years). Regarding each age group, voice samples were recorded from eight children ($M$ = 9.1; $SD$ = 0.4), eleven adolescents ($M$ = 15.1; $SD$ = 1.3), twenty-four young adults ($M$ = 21.5; $SD$ = 2.82), five adults ($M$ = 48.4; $SD$ = 3.1), and five older adults ($M$ = 63.6; $SD$ = 2.6). No inclusion criteria was used, only the participant's age.

**Procedure**

***Voice Recording.*** Every participant from each age group allowed the recording of their voices, signing the informed consent. Children and adolescents were previously authorized by their parents, through written consent, to participate in this study.

Recordings were made in a quiet room, in different places according to the age group (e.g. primary school, senior university) with an Edirol R-09 recorder and Shure PG48 microphone. Voice recordings from young adults were performed in an anechoic chamber. This was due to participant's availability, which young adults were the only age group available to perform the recording in Minho University.

Vocal samples expressing four positive (achievement/triumph, pleasure, relief, amusement) and four negative emotions (sadness, disgust, anger, fear) were recorded, plus surprise and a neutral expression. These categories were selected based on previous studies (Sauter et al., 2010).

Participants were instructed to utter emotional sounds without any verbal content, including interjections, and to be relatively brief. At the same time, participants were instructed to try to be as spontaneous and natural as possible (Lima et al., 2013) and to vocalize more than one vocalization per emotion, with high and low intensity, to ensure some variability.

After the instructions, vocal expressions were elicited using three types of approaches. Initially, participants watched videos that aimed to elicit a specific emotion, and afterwards were asked to utter a congruent sound with the video context (e.g., in a video aiming to elicit disgust, participants saw someone vomiting). Then, participants read aloud some sentences that represented specific emotional scenarios, based on Lima et al. (2013), and were again asked to utter a congruent sound with the intended emotion. Lastly, participants were asked to produce vocal expressions of each emotional category using the vowel 'a'.

For children, adapted videos and sentences were selected, in order to facilitate performance in the task. Also, instructions were more but simplified due to facilitate comprehension of the task. PowerPoint was used to run the instructions. The session lasted between 20-25 minutes for every age group.

***Voice Editing.*** After the voice recordings, physiological sounds were excluded, such as sneezing or a hiccup, as well as verbal interjections known as "affect emblems" (Schroder, 2003), and sounds with recording problems (e.g., with background noise). Voice recordings were edited based on the following steps. Initially, the background noise was removed using version 2.1.1 of Audacity software (®) (Audacity Team, 2014), followed by intensity normalization (i.e., 70 dB) using Praat software version 5.1.05 (Boersma & Weenik, 2009). After this, sounds were segmented individually utilizing Praat (Boersma & Weenik, 2009). These steps aimed to ensure all stimuli had the appropriate acoustic parameters for the next phase of this study.

***Selection of Vocal Samples.*** A group of three researchers chose the best vocal samples from four speakers for each age group, representing a total of 20 speakers (children: $M = 9$, $SD = 0$; adolescents: $M = 14.25$, $SD = 0.5$; young adults: $M = 21.25$, $SD = 0.96$; adults: $M = 49.25$, $SD = 2.9$; elderly: $M = 64.25$, $SD = 2.5$),divided equally by sex and age group. Each judge heard every vocalization uttered by each participant of the study, and individually selected the sound he/she judged to represent the best portrayal of each emotion category. The vocalizations that gathered consensus were immediately selected. Those vocal samples that

were not consensual were discussed between the judges who then reached a consensus regarding the better sound that characterized a specific emotion. In the end, 200 vocalizations were selected for validation (10emotion*5age groups*4speakers = 200 vocalizations), which were divided equally in two experimental sessions. These vocalizations were produced by five age groups, and each age group had 4 speakers (i.e., 2 male and 2 female).

**Stage 2: Ratings of Vocalizations**

**Participants.** A group of 48 young adults (24 females, 24 males, mean age = 22.4 years, *SD* = 2.2) and 36 children (18 females, 18 males, mean age = 8.7 years, *SD* = 1) participated in the second stage in which the previously recorded vocalizations were evaluated. All participants were native speakers of European Portuguese, did not have formal musical training, and had normal auditory acuity (i.e., pitch discrimination). No psychiatric or neurological illness were reported in sociodemographic questionnaire, including head trauma and substance abuse.

**Participants' Assessment.** After signing an informed consent form, participants filled out a relevant sociodemographic questionnaire (i.e., sex, age, education, occupation, musical training, psychiatric or neurological history, socioeconomic level). In the case of children, permission to participate in the study was also obtained from their parents, through signed informed consent.

**Procedure.** The experiment was divided in two sessions. Each session comprised categorical and valence assessment tasks (pseudo-randomized), and, in each task, 100 vocalizations were assessed. Ten vocalizations per emotion were presented in each task, divided equally by age group and sex. Instructions with an illustrative example of a trial, and a training phase preceded the experimental phase. In this phase, participants had access to a sheet of paper with different scenarios in order to contextualize some of the emotions (Sauter et al., 2007) (same scenarios used in the recording sessions).

The training phase followed the same procedure of the experimental phase, but included only four vocalizations selected from a validated database (Lima et al., 2003). For better comprehension, participants were told that every sound represented a specific emotion or absence of emotion, and they had to press the number associated with the intended emotion (e.g. 1 represents «sadness»). This training occurred in order to ensure familiarization with the task and with the instructions. For children, instructions and scenarios were adapted. This

happened in order to allow a greater comprehension of the task and to reduce fatigue. Therefore, two sessions were administrated, with each including also 100 vocalizations.

The categorization assessment was a forced choice task. After stimulus presentation, participants had to choose between the following options: sadness, fear, disgust, anger, surprise, relief, pleasure, amusement, achievement and neutrality (Figure 1). After each stimulus, participants had the chance to replay the vocalization heard, the number of times necessary for its understanding. Every label was visible after stimulus presentation, with an emoticon representing each emotion to facilitate emotional identification. The response options ranged from zero to nine (0-9) and were presented in a specific order (i.e., 1-4: negative emotions, 5: surprise, 6-9: positive emotions, 0: neutral; Figure 2).

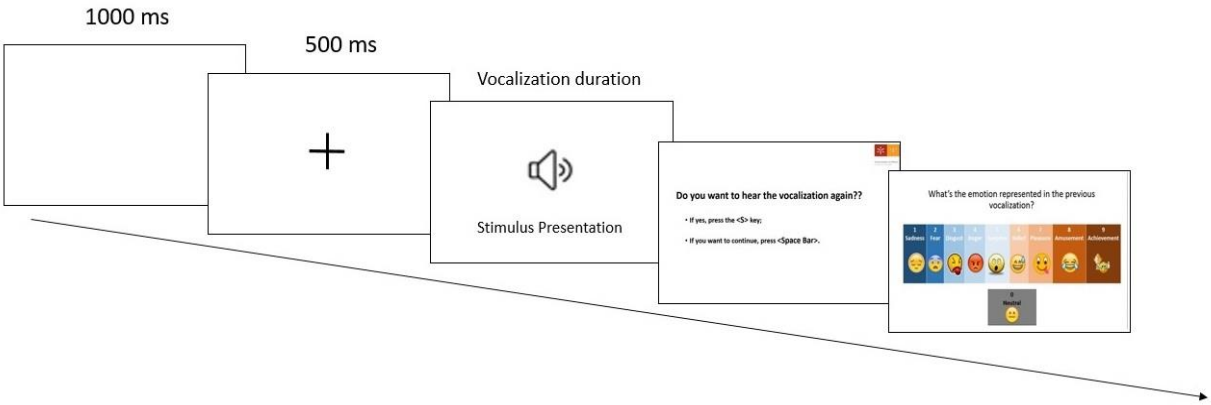


*Figure 1.* Illustration of an experimental trial of the categorization task

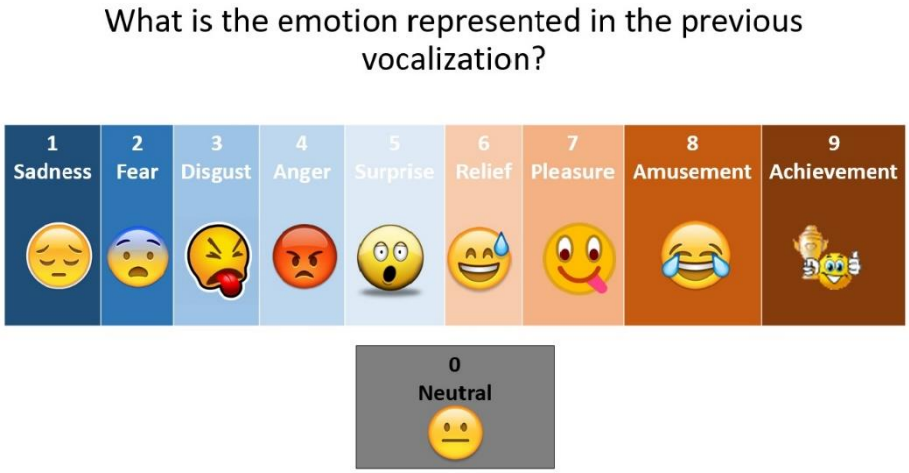


*Figure 2.* Response options in the categorization task

The valence assessment had an identical layout as the categorization assessment (Figure 3). The only difference occurred after stimulus presentation, where each participant was asked to rate valence scale (Figure 3). Valence pretends to assess how pleasant the stimulus is (Russell, 1980). The following scale varied from one to nine (Bradley & Lang, 1994), namely from "very unpleasant" to "very pleasant".

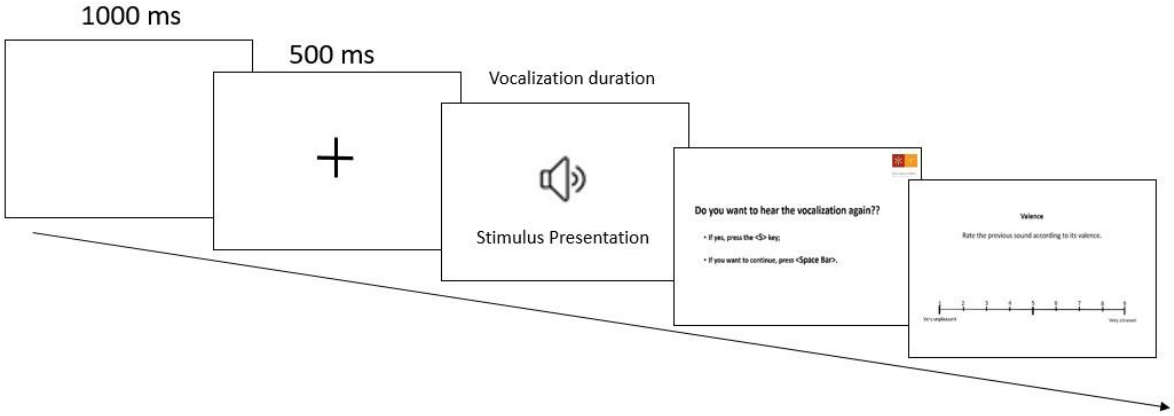


*Figure 3.* Illustration of an experimental trial of the valence assessment

In both assessments, vocalizations were presented in a random order, trough headphones (Sony MDR-NC6) and no feedback was provided in both training and experimental blocks. These tasks were run in Superlab version 5.0 (Cedrus Corporation, 2014).

## Results

For the statistical analyses, the IBM SPSS Statistics 23.00 (IBM Corp. Released, 2015) software package was used. Only significant results are presented ($p < 0.05$). No outliers were found in categorization and valence assessment, to do so it was observed if any participant had hit rate/ mean valence rate above or below three standard deviations according to the overall mean.

In order to probe that every emotion modality was recognized above chance a one sample *t* test was computed. Next, a mixed analysis of variance (i.e., ANOVA) was computed for categorical and valence ratings separately. Mixed ANOVA was computed with emotional category, speaker's sex and speaker's age as within-subjects, and participant's sex and participant's age as between-subjects factor. Hit rate was calculated through the number of correct responses according to emotional category, divided by the total number of

presentations. When sphericity was violated in this statistic test, the Greenhouse-Geisser adjustment was used. Due to the high number of significant interaction and spatial restrictions, the Results section focuses on main effects and interactions pertinent for the purpose of this study.

**Categorization Task**

Consistency among each group was verified by a high inter-participant reliability in categorizations for young adults (Cronbach's $\alpha = .825$) and for children (Cronbach's $\alpha = .821$). One sample *t* test was run on participant ratings for all emotion categories, in order to probe that every emotion modality was recognized above chance.

In young adults, achievement, $t(47) = 16.893$, anger, $t(47) = 33.045$, disgust, $t(47) = 125.686$, amusement, $t(47) = 55.639$, fear, $t(47) = 16.950$, neutral, $t(47) = 25.298$, pleasure, $t(47) = 35.218$, relief, $t(47) = 32.147$, sadness, $t(47) = 29.676$, and surprise, $t(47) = 25.858$, were recognized above chance, all emotions were significant after Bonferroni correction, $p < .001$. However, several patterns of errors were observed in young adults (Table 1). Fear vocalizations were confused with surprise vocalizations and vice versa, while anger vocalizations were confused with disgust and achievement vocalizations were confused with amusement and anger.

In children, all emotions were also recognized above chance: achievement, $t(21) = 9.293$, anger, $t(21) = 19.956$, disgust, $t(21) = 30.874$, amusement, $t(21) = 30.432$, fear, $t(21) = 9.327$, neutral, $t(21) = 13.051$, pleasure, $t(21) = 19.952$, relief, $t(21) = 15.760$, sadness, $t(21) = 20.475$, and surprise, $t(21) = 19.186$, all emotions were significant after Bonferroni correction, $p < .001$. In this age group, patterns of errors were observed more often (Table 2). Most emotions were confused with neutral expression, which may indicate that, when in doubt, participants of this age group decided for this option.

First of all, the emotions that presented the biggest recognition accuracy were not confused with any other emotion, namely disgust, amusement and pleasure. Regarding the rest of the categories, all of them were confused with neutral vocalizations. Other effects worthy of mention are the confusion between achievement and amusement, between anger and both surprise and disgust, between surprise and both fear and relief, and finally the neutral and relief emotions were only confused with surprise. Further, fear vocalizations,

besides being confused with neutral expressions, as stated before, were categorized more often as surprise.

So, some vocalizations had a recognition accuracy below chance (i.e., 10%). In young adults, the vocalizations with this low hit rate totaled 12 stimuli, namely, two in achievement, three in anger, three in fear, three in sadness and one in surprise. Regarding children, there was a higher number of vocalizations with a hit rate below chance with the total of 17 stimuli, namely, six in achievement, two in anger, one in disgust, four in fear, one in relief, and three in sadness.

A main effect of emotional category was observed, $F(5.98,394.85) = 89.73$, $p < .001$, $\eta_p^2 = .58$. Pairwise comparisons analyses revealed that amusement presented a significant higher accuracy in comparison with achievement, $p < .001$, anger, $p < .001$, fear, $p < .001$, sadness, $p < .001$, surprise, $p < .001$, and relief, $p = .012$. Disgust followed a similar pattern, and was significantly better recognized than achievement, anger, fear, sadness, and surprise, all comparisons were significant with $p < .001$. On the other hand, achievement vocalizations were significantly less accurately recognized than the previously mentioned emotional categories (i.e., disgust and amusement) and neutral, $p < .001$, pleasure, $p < .001$, and relief, $p = .012$. At the same time, fear vocalizations were also less accurately recognized than neutral expressions, $p < .001$, pleasure, $p < .001$, and relief, $p < .002$.

Moreover, a main effect of speaker's age was observed, $F(4,264) = 72.93$, $p < .001$, $\eta_p^2 = .59$. Subsequent pairwise comparisons showed that vocalizations uttered by older adult speakers were less accurately recognized than those conveyed by children, adolescents, young adults, and adults (all were significant - $p < .001$). Furthermore, vocalizations uttered by children were better recognized than those uttered by adolescents, young adults, adults and elderly (all were significant - $p < .001$).

**Table 1.** Percentage of responses for each emotion category in young adults

| Response | Intended emotion | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Achievement | Anger | Disgust | Amusement | Fear | Neutral | Pleasure | Relief | Sadness | Surprise |
| Achievement | **45.9 (25.3)** | 0.3 | 0.4 | 1.5 | 0.8 | 0.3 | 1.3 | 0.2 | 0.4 | 1.0 |
| Anger | 11.1 | **61.0 (31.8)** | 1.5 | 0.1 | 2.1 | 1.1 | 1.7 | 1.0 | 1.1 | 0.8 |
| Disgust | 3.3 | 15.3 | **89.1 (22.8)** | 0.0 | 3.1 | 0.9 | 1.1 | 1.3 | 3.5 | 3.1 |
| Amusement | 11.3 | 0.2 | 0.3 | **92.9 (5.7)** | 0.5 | 0.1 | 1.1 | 0.0 | 5.5 | 0.7 |
| Fear | 3.4 | 3.9 | 0.9 | 0.2 | **45.2 (24.8)** | 1.3 | 0.3 | 0.3 | 5.1 | 21.7 |
| Neutral | 7.7 | 6.1 | 3.5 | 1.3 | 15.3 | **80.5 (16.5)** | 2.4 | 5.2 | 7.8 | 4.0 |
| Pleasure | 4.5 | 3.4 | 3.0 | 2.7 | 0.1 | 1.3 | **76.4 (21.1)** | 11.9 | 4.4 | 3.5 |
| Relief | 6.9 | 5.7 | 0.4 | 0.5 | 3.2 | 6.1 | 9.3 | **68.6 (23.3)** | 8.9 | 8.1 |
| Sadness | 0.5 | 0.5 | 0.2 | 0.6 | 1.6 | 0.8 | 0.4 | 2.8 | **58.3 (30.5)** | 0.2 |
| Surprise | 5.3 | 3.4 | 0.6 | 0.2 | 28.0 | 7.5 | 6.0 | 8.6 | 4.9 | **56.8 (26.0)** |
| Total | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

*Note.* Bold cells represent accurate categorizations with underlying standard deviation

**Table 2.** Percentage of responses for each emotion category in children

| Response | Intended emotion | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Achievement | Anger | Disgust | Amusement | Fear | Neutral | Pleasure | Relief | Sadness | Surprise |
| Achievement | **34.3 (21.5)** | 1.4 | 1.1 | 7.0 | 2.3 | 1.6 | 0.9 | 1.1 | 2.0 | 2.5 |
| Anger | 7.3 | **46.4 (31.1)** | 3.4 | 0.2 | 2.5 | 0.9 | 3.9 | 3.4 | 2.0 | 0.9 |
| Disgust | 3.4 | 10.5 | **79.1 (26.0)** | 0.2 | 2.5 | 5.0 | 4.5 | 1.4 | 4.5 | 2.3 |
| Amusement | 20.2 | 0.7 | 0.7 | **83.0 (6.9)** | 2.7 | 0,0 | 1.8 | 0.9 | 2.3 | 0.2 |
| Fear | 3.0 | 6.4 | 3.0 | 0.2 | **29.8 (17.4)** | 3.0 | 0.7 | 2.7 | 4.8 | 13.0 |
| Neutral | 10.7 | 14.5 | 5.7 | 1.4 | 17.5 | **65.7 (9.7)** | 7.3 | 10.5 | 16.6 | 15.0 |
| Pleasure | 4.8 | 3.0 | 2.7 | 2.7 | 1.1 | 3.6 | **70.2 (20.3)** | 6.6 | 3.9 | 2.0 |
| Relief | 6.6 | 5.9 | 1.1 | 2.3 | 6.1 | 4.5 | 7.3 | **59.5 (21.2)** | 5.2 | 14.5 |
| Sadness | 1.4 | 0.5 | 0.2 | 2.0 | 1.6 | 0.9 | 0.0 | 1.6 | **52.5 (32.8)** | 0.7 |
| Surprise | 8.4 | 10.9 | 3.0 | 0.9 | 33.9 | 14.8 | 3.4 | 12.3 | 6.1 | **48.9 (14.6)** |
| Total | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

A main effect of speaker's sex was also revealed, $F(1,66) = 49.81$, $p < .001$, $\eta_p^2 = .43$. Pairwise comparisons revealed that vocalizations uttered by female speakers were associated with higher accuracy rates in comparison with those conveyed by male speakers, $p < .001$.

Furthermore, a significant interaction between speaker's sex and participant's age was also shown, $F(1,66) = 10.73$, $p = .002$, $\eta_p^2 = .14$. Pairwise comparisons revealed that, for both children and adolescent listeners, vocalizations uttered by female speakers were better recognized than when uttered by male speakers (both were significant - $p < .001$). In addition, a significant interaction between emotion and speaker's sex was shown, $F(9,100) = 5.99$, $p < .001$, $\eta_p^2 = .25$. Pairwise comparisons revealed significant differences in hit rates when vocalizations were produced by male *vs.* female speakers: in comparison with male vocalizations, vocalizations uttered by female speakers were associated with higher hit rates for the following specific emotional categories: achievement, anger, fear, relief and sadness (all comparisons were significant $p < .001$).

A significant interaction between speaker's age and speaker's sex was revealed, $F(3.58,397.85) = 15.77$, $p < .001$, $\eta_p^2 = .19$. Pairwise comparisons showed that, when vocalizations were uttered by female speakers, subjects had a higher accuracy rate in recognizing emotions vocalized by young adults, $p < .001$, adults, $p < .001$, and elderly, $p = .003$. No differences were found in speaker's sex when vocalizations were uttered by children and adolescents ($p > .05$).

At last, a significant interaction was observed between speaker's age and participant's age, $F(4,264) = 2.51$, $p = .043$, $\eta_p^2 = .04$. Pairwise comparisons showed that, young adults had a better recognition accuracy in vocalizations when were uttered by children, in comparison with the ones conveyed by adolescents, young adults, adults, and elderly, all comparisons were significant $p < .001$; while vocalizations uttered by elderly had the lowest accuracy comparing to stimuli conveyed by children, adolescents, young adults, and adults, all comparisons were significant with $p < .001$. Regarding children participants, the same pattern was observed, vocalizations uttered by children had the highest recognition accuracy, in comparison with stimuli conveyed by adolescents, $p = .011$, young adults, $p = .001$, adults, $p = .001$, and elderly, $p < .001$; on the contrary vocalizations uttered by elderly had the lowest recognition accuracy comparing to stimuli conveyed by children, adolescents, young adults, and adults, all comparisons were significant $p < .001$.

Regarding differences in between-subjects factor, a main effect in participant's age was observed, $F(1,66) = 41.19$, $p < .001$. Pairwise comparison revealed a higher recognition accuracy in young adults group in comparison with children regarding overall hit rate, $p < .001$. Moreover, no sex differences were found between participants ($p > .05$).

**Valence Assessment**

Consistency among each group was demonstrated by a high inter-participant reliability in valence assessments for both young adults (Cronbach's $\alpha = .921$) and for children (Cronbach's $\alpha = .879$) participants.

A main effect of emotional category was revealed, $F(9,594) = 11.19$, $p < .001$, $\eta_p^2 = .81$. Multiple pairwise comparisons revealed that amusement was perceived as more pleasant than the other vocal emotional categories (all comparisons were significant $p < .005$). Pleasure and achievement were also judged as more pleasant than the other emotional categories ($p < .001$), except amusement. Relief were rated as more pleasant than all negative emotions (i.e., anger, disgust, sadness and fear), neutral and surprise, all comparisons were significant $p < .005$. On the other hand, sadness and disgust vocalizations were significantly rated as less pleasant than every other emotional category, all comparisons were significant $p < .005$, except anger ($p > .05$). Anger vocalizations had lower valence ratings than all other emotional categories (all comparisons were significant $p < .005$), with the exception of sadness, disgust and fear. At last, fear vocalizations only were rated as more unpleasant than achievement, amusement, relief, pleasure, neutral and surprise, all comparisons were significant $p < .005$. Regarding neutral and surprise vocalizations, significant differences were observed relative the other emotional categories, with the exception of fear, $p > .05$. These emotion categories had higher valence ratings than negative emotions, (i.e., sadness, disgust and anger), and lower ratings than positive emotions (i.e., achievement, amusement, relief and pleasure), all comparisons were significant $p < .005$.

A main effect of speaker's sex was observed $F(1,66) = 14.84$, $p < .001$, $\eta_p^2 = .18$. Pairwise comparisons revealed that vocalizations uttered by male speakers were rated as more pleasant, $p < .001$. A significant interaction was also revealed between speaker's sex and emotional category, $F(6.93,457.26) = 16.04$, $p < .001$, $\eta_p^2 = .20$. Multiple pairwise comparisons revealed that, compared to female vocalizations, male vocalizations were rated as more pleasant for the following emotional categories: disgust, $p < .001$, amusement, $p =$

.018, fear, $p = .01$, pleasure, $p < .001$, and sadness, $p < .001$. On the other hand, achievement vocalizations were rated as more pleasant when uttered by female relative to male speakers, $p = .05$.

A main effect of speaker's age was not observed ($p > .05$): However a significant interaction between speaker's age and emotional category was revealed, $F(36,2376) = 12.97$, $p < .001$, $\eta_p^2 = .16$. Multiple pairwise comparisons revealed the following specific effects:

For achievement vocalizations, the lowest valence rates were obtained when the stimuli were uttered by older adult speakers in comparison with the other age groups, with the contrary effect observed in children group, all comparisons were significant $p < .001$. Regarding anger vocalizations, when uttered by older adult speakers, they were assessed as more pleasant than those produce by children, $p = .001$. When adults conveyed disgust vocalizations, they were perceived as more pleasant compared to all the other age groups, $p < .001$. For amusement, vocalizations produced by adults were also perceived as more pleasant, but only when compared with children, $p < .001$, and elderly, $p = .01$. Regarding neutral vocalizations, when conveyed by children, a higher valence rate was observed when compared with adult, $p = .006$, and young adult speakers, $p = .001$. When pleasure vocalizations were uttered by older adult speakers, valence ratings were higher than when produced by adults, $p = .006$, adolescents, $p = .001$, and young adults, $p < .001$. Regarding relief, vocalizations uttered by older adults had higher valence ratings than when uttered by adolescents, $p = .008$. Regarding sadness, when vocalizations were uttered by young adult speakers, lower valence ratings were observed when compared with adolescents, $p = .019$, adults, $p < .001$, and older adults, $p < .001$. Valence ratings of surprise and fear vocalizations showed no differences as a function of speaker's age ($p > .05$).

Furthermore, a significant interaction between participant's age and emotional category was revealed, $F(5.22,344.22) = 5.24$, $p < .001$, $\eta_p^2 = .07$. Pairwise comparisons revealed that young adult's listeners rated anger, $p = .003$, disgust, $p = .009$, fear, $p = .001$, and surprise, $p = .027$ as more unpleasant in comparison with children listeners.

Regarding between subjects analysis, no differences in valence rates in participant's age and participant's sex ($p > .05$), however significant interactions of interest with these variables are reported above.

## Discussion

This study allowed us to study a variety of emotional categories and distinct characteristics of the speaker, namely age (i.e., children, adolescents, young adults, adults and elderly) and sex, as well as the effects of listener's characteristics (age and sex) on vocal emotional recognition and affective ratings of vocalizations. To the best of our knowledge, this is the first study to examine differences in vocal emotional processing as a function of speaker's age, which allowed us to probe if vocal emotional recognition and the assessment of the pleasantness of a vocalization vary when a subject interacts with persons of different ages through the life span. Besides that, it also allowed us to test whether and how participant's age (children and young adults) and sex modulate the decoding of emotional information from vocal cues.

Our results confirmed our initial prediction that both speaker's age and speaker's participant affect vocal emotional processing.

Emotional vocalizations were associated with differences in recognition, but not in valence assessment, as a function of speaker's age. When vocal emotions were conveyed by older adult speakers, lower recognition accuracy was observed in comparison with the other age groups, whereas vocalizations uttered by the youngest speakers (children from 8 to 10 years) were the most accurately recognized. The low recognition of vocal emotions produced by older adults suggests is consistent with reports of decreased emotional recognition as a function of aging, proposed by a neuropsychological approach (Mill et al., 2009; Ruffman et al., 2008), but this finding may also indicate that the ability to express emotions is decreased as a function of aging. The opposite finding (i.e., facilitated recognition of vocal emotions produced by children) suggests that children have an advantage communicating vocal emotions, which may be due to the need of care and protection from the surrounding environment (Trainor, Austin, & Desjardins, 2000). However, further studies are needed to understand if this advantage is associated with differences in vocal expression of emotions in younger ages (e.g., pitch differences that may facilitate emotional decoding), or whether in general human beings are tuned to detect vocal emotions expressed by subjects who are inherently associated with higher needs for protection and care (Babchuk et al., 1985).

Furthermore, it was hypothesized that vocalizations conveyed by speakers of the same age group of participants would have better accuracy. This was not observed, because both groups (i.e., children and young adults), followed the same pattern. Vocalizations conveyed

by children were better decoded, and on the contrary the ones uttered by elderly had the lowest recognition accuracy. This emphasizes again the advantage of children in expressing emotions, through different age groups.

Speaker's sex was also an important variable that modulated vocal emotional recognition, following previous studies (Belin et al., 2008). Regarding recognition accuracy, participants were consistently better, in both participant's age groups, in decoding vocal emotions expressed by female speakers, as hypothesized and in line with previous studies (Belin et al., 2008). However, no interaction was observed between speaker's sex and participant's sex in terms of accuracy, contrary to what was observed by Belin et al. (2008).

The results obtained in this study also revealed important effects of participants' characteristics, which should be accounted for by studies on vocal emotional processing. Overall, vocalizations were better recognized by young adult listeners in comparison with children (Phillips et al., 2002; Tonks et al., 2007), as hypothesized. The decreased emotional recognition abilities in children can be explained by maturational effects: as the brain is still evolving, and particularly prefrontal regions (associated with higher-order cognition) are not fully developed, they are less capable to evaluate the emotional content of stimuli, which can lead to a high pattern of errors and lower hit rates as observed in the present study. This study do not showed differences between sexes in emotional recognition regarding non-verbal vocalizations, in line with Lima et al. (2013) and Sauter et al. (2013), but not predicted by our hypotheses (Belin et al., 2008; Schirmer et al., 2002).

Another major finding from this study was the observation that vocalizations were rated as more pleasant by children than by young adult participants, in specific emotions. This tendency to evaluate stimuli as more positive – specifically, anger, disgust, fear and surprise - may be related to developmental effects in emotional processing. Specifically, this pattern of findings may suggest that the proper notion of danger in the surrounding environment is not yet fully developed in children. This may happen due to the fact that the avoidance system (i.e., a system that elicits avoidance responses from threatening stimuli) is not yet fully developed until 10 years of age (Lang & Bradley, 2013). On the other hand, when comparing valence ratings regarding speaker's sex, vocalizations uttered by male speakers had higher rates, namely, in amusement, pleasure, fear, disgust, and sadness. Specifically, they may consider these negative stimuli as more unpleasant, due to the fact that female are believed to be more emotional expressive (Belin et al., 2008). However, in positive emotions previously

mentioned, the contrary was observed, emphasizing the need of further investigation in this topic.

Besides showing the effects of both speaker's and participant's characteristics on vocal emotional processing, this study also allowed the validation of a corpus comprising non-verbal vocalizations and conveyed by speakers with distinct ages, divided equally by sex. This set comprises ten emotional categories: achievement, amusement, relief, pleasure, surprise, fear, anger, disgust, sadness and neutral. Only vocalizations with higher rates than chance (i.e., higher than 10%) were selected for the current analysis. So, the final corpus includes 188 vocalizations validated by young adults, and 183 validated by children. Information about the number of selected stimuli is described in Table 3 and 4.

The vocalizations included in this corpus were associated with a relatively high mean accuracy rate ($M$ = 73% for young adults, $M$ = 61% for children). However, including those vocalizations that were withdrawn from the study due to above chance accuracy, this recognition rate becomes slightly lower ($M$ = 67% for young adults, $M$ = 57% for children). The overall recognition rate is slightly lower than previous studies (Belin et al., 2008; Lima et al., 2013; Sauter et al., 2013), which can be due to the fact that more diverse vocalizations (from participants with different ages and, plausibly, less effective in conveying emotions) were included in the current set.

Specifically, this study showed that amusement and disgust were the most accurately recognized emotions, whereas achievement and fear were the least recognized regarding both groups, in line with previous studies (Belin et al., 2008; Lima et al., 2013; Sauter et al., 2013). Still, all emotional categories were consistently associated with the intended emotion. In line with Lima et al. (2013), fear vocalizations were often confused with surprise, and vice versa, as hypothesized, which may be related to the fact that these emotions have similar meaning and at the same time share similar acoustic patterns. Furthermore, anger vocalizations were confused with disgust, whereas achievement vocalizations were confused with amusement and anger, in line with previous studies (Lima et al., 2013). Regarding valence affective ratings, the emotional categories that had higher recognition rates were also the ones with lowest and highest valence rates, respectively: disgust and amusement.

**Table 3.** Number of vocalizations validated with young adult participants, per speaker's age and sex

| | Children | | Adolescents | | Young Adults | | Adults | | Elderly | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Female | Male | Female | Male | Female | Male | Female | Male | Female | Male | |
| Achievement | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 18 |
| Anger | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 17 |
| Disgust | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 20 |
| Amusement | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 20 |
| Fear | 2 | 2 | 2 | 2 | 2 | - | 2 | 2 | 1 | 2 | 17 |
| Neutral | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 20 |
| Pleasure | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 20 |
| Relief | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 20 |
| Sadness | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 17 |
| Surprise | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 19 |
| Total | 19 | 20 | 19 | 20 | 20 | 18 | 20 | 19 | 16 | 17 | 188 |

**Table 4.** Number of vocalizations validated with children participants, per speaker's age and sex

| | Children | | Adolescents | | Young Adults | | Adults | | Elderly | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Female | Male | Female | Male | Female | Male | Female | Male | Female | Male | |
| Achievement | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 2 | - | 1 | 14 |
| Anger | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 18 |
| Disgust | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 19 |
| Amusement | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 20 |
| Fear | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 16 |
| Neutral | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 20 |
| Pleasure | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 20 |
| Relief | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 19 |
| Sadness | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 17 |
| Surprise | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 20 |
| Total | 19 | 20 | 18 | 19 | 20 | 17 | 20 | 19 | 14 | 17 | 183 |

However, some limitations of the present study need to be highlighted. Recordings were made in different locations, according to participant's availability, which may have cause variability in the acoustic quality of the sounds. Also, the duration of the task may have increased fatigue and boredom, which may have affected participants' performance. Also, as we have only tested children and young adult participants, it is pertinent to examine if similar effects are observed with participants from other age groups (e.g., adolescent and older adult listeners), for a more comprehensive view of vocal emotional recognition across the life span.

## Conclusions

The present study points to an effect of both speaker's and participant's features (age and sex) on vocal emotional processing. Regarding speaker's sex, a higher accuracy was observed for vocal emotions uttered by female speakers. On the other hand, vocal emotions uttered by older adult speakers were the least accurately decoded, while the highest accuracy was obtained for vocalizations uttered by children.

This corpus is innovative as it provides vocal samples from speakers of different ages, and for 10 emotional categories. It provides controlled and validated stimuli that can be used in the general population, including in research with children. To the best of our knowledge, this is the first database comprising vocalizations uttered by five distinct age groups (i.e., children, adolescents, young adults, adults, and elderly).

This set may be used for research or, in the future, for clinical purposes, namely in experimental studies of vocal emotional processing, as well as in neuropsychological assessments. Future studies may test whether vocal emotions produced by different age groups elicit different brain responses (e.g., through neuroimaging) or differences in the time course of vocal processing (e.g., through event related potentials). Additionally, acoustic analyses of these stimuli are also pertinent, in order to explore the effects of acoustic features of the vocalizations on vocal emotional processing across the life span.

## References

Amilon, K., van de Weijer, J., & Schötz, S. (2007). The Impact of Visual and Auditory Cues in Age Estimation Kajsa Amilon, Joost van de Weijer, Susanne Schötz. *Speaker Classification II*, 10–21. http://doi.org/10.1007/978-3-540-74122-0

Audacity Team (2014). *Audacity(®): Free Audio Editor and Recorder* (Version 2.1.1). Retrieved April 20, 2014, from http://audacity.sourceforge.net/

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.

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice : neural correlates of voice perception. *Trends in Cognitive Sciences*, *8*(3), 129–135. http://doi.org/10.1016/j.tics.2004.01.008

Belin, P., Filllion-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. http://doi.org/10.3758/BRM.40.2.531

Bianchin, M., & Angrilli, A. (2012). Physiology & Behavior Gender differences in emotional responses : A psychophysiological study. *Physiology & Behavior*, *105*(4), 925–932. http://doi.org/10.1016/j.physbeh.2011.10.031

Boersma, P., & Weenink, D. (2009). *Praat: doing phonetics by computer* (Version 5.1.05). Retrieved May 1, 2009, from http://www.praat.org/

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(1)*, 49-59.

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. http://doi.org/10.3758/BRM.42.1.74

Cedrus Corporation. (2014). Superlab (Version 5.0) [computer software]. San Pedro: Cedrus. Available, from http://www.superlab.com/

Ekman, P. (1992). An argument for basic emotions. In *Cognition and Emotion* (pp. 169–200).

Fujimura, T., Matsuda, Y., Katahira, K., Okada, M., & Okanoya, K. (2016). Categorical and dimensional perceptions in decoding emotional facial expressions. *Cognition and Emotion*, *26(4)*, 587–601. http://doi.org/10.1080/02699931.2011.595391

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*(6), 401–416. http://doi.org/10.1016/j.evolhumbehav.2006.05.002

Hawk, S. T., Kleef, G. A. Van, Fischer, A. H., & Schalk, J. Van Der. (2009). " Worth a Thousand Words ": Absolute and Relative Decoding of Nonlinguistic Affect Vocalizations, *9*(3), 293–305. http://doi.org/10.1037/a0015178

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. http://doi.org/10.1037/0033-2909.129.5.770

Lang, P. J., & Bradley, M. M. (2013). Appetitive and Defensive Motivation: Goal-Directed or Goal-Determined? *Emotion Review : Journal of the International Society for Research on Emotion, 5(3)*, 230–234.

Laukka, P., Juslin, P., & Bresin, R. (2005). A dimensional approach to vocal expression of

emotion. *Cognition and Emotion*, *19*(5), 633–653. http://doi.org/10.1080/02699930441000445

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*(4), 1234–1245. http://doi.org/10.3758/s13428-013-0324-3

Mill, A., Allik, J., Realo, A., & Valk, R. (2009). Age-Related Differences in Emotion Recognition Ability : A Cross-Sectional Study. *Emotion*, *9*(5), 619–630. http://doi.org/10.1037/a0016562

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.

Owren, M. J., Amoss, R. T., & Rendall, D. (2011). Two Organizing Principles of Vocal Production : Implications for Nonhuman and Human Primates. *American Journal of Primatology*, *73*(6), 530–544. http://doi.org/10.1002/ajp.20913

Paulmann, S., Pell, M. D., & Kotz, S. A. (2008). How aging affects the recognition of emotional speech. *Brain and Language*, *104*(3), 262–269. http://doi.org/10.1016/j.bandl.2007.03.002

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. http://doi.org/10.1016/j.biopsycho.2015.08.008

Phillips, L. H., MacLean, R. D. J., & Allen, R. (2002). Age and the understanding of emotions: neuropsychological and sociocognitive perspectives. *The Journals of*

*Gerontology. Series B, Psychological Sciences and Social Sciences*, *57*(6), P526–P530. http://doi.org/10.1093/geronb/57.6.P526

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect : An integrative approach to affective neuroscience , cognitive development , and psychopathology. *Development and Psychopathology*, *17*, 715–734.

Ruffman, T., Henry, J. D., Livingstone, V., & Phillips, L. H. (2008). A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. *Neuroscience and Biobehavioral Reviews*, *32*(4), 863–881. http://doi.org/10.1016/j.neubiorev.2008.01.001

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*, 1161–1178.

Sauter, D. A. (2007). *An investigation into vocal expressions of emotions: The roles of valence, cultures and acoustic factors*. University of London. Retrieved from http://brainb.psyc.bbk.ac.uk/PDF/thesis.pdf

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. http://doi.org/10.1080/17470211003721642

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

Sauter, D. A., Panattoni, C., & Happ, F. (2013). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology*, *31*(1), 97 – 113. http://doi.org/10.1111/j.2044-835X.2012.02081.x

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2002). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research*, *14*(2), 228–233.

Schroder, M. (2003). Experimental study of affect bursts. *Speech Communication*, *40*(1), 99–116. http://doi.org/10.1016/S0167-6393(02)00078-X

Stevens, J. S., Hamann, S. B., Stevens, J. S., & Hamann, S. (2012). Sex differences in brain activation to emotional stimuli : A meta-analysis of neuroimaging studies neuroimaging studies. *Neuropsychologia*, *50*(7), 1578–1593. http://doi.org/10.1016/j.neuropsychologia.2012.03.011

Tonks, J., Williams, W. H. U. W., Frampton, I. A. N., Yates, P., & Slater, A. (2007). Assessing emotion recognition in 9 – 15-years olds : Preliminary analysis of abilities in reading emotion from faces , voices and eyes. *Brain Injury*, *21*(6), 623–629. http://doi.org/10.1080/02699050701426865

Trainor, L. J., Austin, C. M., & Desjardins, N. (2000). Is Infant-directed speech prosody a result of the vocal expression of emotion? *Psychological Science*, *11*(3), 188–196.

Turkstra, L. S. (2000). Should my shirt be tucked in or left out ? The communication context of adolescence. *Aphasiology*, *14*(4), 349 – 364. http://doi.org/10.1080/026870300401405

Wiethoff, K., Wildgruber, D., & Becker, H. (2008). Cerebral processing of emotional prosody — Influence of acoustic parameters and arousal parameters and arousal. *Neuroimage*, *39*(2), 885–893. http://doi.org/10.1016/j.neuroimage.2007.09.028