Changes in Vocal Emotion Recognition Across the Life Span

Maria Amorim, Andrey Anikin, Augusto J. Mendes, César F. Lima, Sonja A. Kotz, and Ana P. Pinheiro


CITATION
Changes in Vocal Emotion Recognition Across the Life Span

Maria Amorim
Universidade de Lisboa

Andrey Anikin
Lund University

Augusto J. Mendes
University of Minho

César F. Lima
Instituto Universitário de Lisboa (ISCTE-IUL)

Sonja A. Kotz
University of Maastricht and Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Ana P. Pinheiro
Universidade de Lisboa and University of Maastricht

The ability to recognize emotions undergoes major developmental changes from infancy to adolescence, peaking in early adulthood, and declining with aging. A life span approach to emotion recognition is lacking in the auditory domain, and it remains unclear how the speaker’s and listener’s ages interact in the context of decoding vocal emotions. Here, we examined age-related differences in vocal emotion recognition from childhood until older adulthood and tested for a potential own-age bias in performance. A total of 164 participants (36 children [7–11 years], 53 adolescents [12–17 years], 48 young adults [20–30 years], 27 older adults [58–82 years]) completed a forced-choice emotion categorization task with nonverbal vocalizations expressing pleasure, relief, achievement, happiness, sadness, disgust, anger, fear, surprise, and neutrality. These vocalizations were produced by 16 speakers, 4 from each age group (children [8–11 years], adolescents [14–16 years], young adults [19–23 years], older adults [60–75 years]). Accuracy in vocal emotion recognition improved from childhood to early adulthood and declined in older adults. Moreover, patterns of improvement and decline differed by emotion category: faster development for pleasure, relief, sadness, and surprise and delayed decline for fear and surprise. Vocal emotions produced by older adults were more difficult to recognize when compared to all other age groups. No evidence for an own-age bias was found, except in children. These findings support effects of both speaker and listener ages on how vocal emotions are decoded and inform current models of vocal emotion perception.

Keywords: emotion, voice, age, speaker, listener

Supplemental materials: http://dx.doi.org/10.1037/emo0000692.supp

The ability to decode emotions from different sensory channels plays an adaptive role early in development and undergoes major developmental changes (Morningstar, Nelson, & Dirks, 2018). Soon after birth, infants are able to discriminate vocal expressions (Grossmann, Oberecker, Koch, & Friederici, 2010). The ability to accurately decode emotions improves from childhood to adulthood (Allgood & Heaton, 2015; Chronaki, Hadwin, Garner, Maurage, & Sonuga-Barke, 2015; Chronaki, Wigelsworth, Pell, & Kotz, 2018; Grosbras, Ross, & Belin, 2018; Matsumoto & Kishimoto, 1983), but declines in older adulthood (Lima, Alves, Scott, & Castro, 2015).
Age-related changes in emotion recognition seem to vary by emotion category (e.g., Chronaki et al., 2015; Thomas, De Bellis, Graham, & LaBar, 2007). For example, there is evidence that the recognition of sadness and happiness in nonverbal vocalizations reaches an adult-like pattern earlier than the recognition of anger and fear (Grosbras et al., 2018). Developmental differences in emotion recognition might also be driven by differences in stimulus type (Grosbras et al., 2018; Tonks, Williams, Frampton, Yates, & Slater, 2007). In the case of facial emotions, inconsistent findings have been reported regarding the age at which adult-level performance is reached (e.g., 11 years of age: Chronaki et al., 2015; Tonks et al., 2007; after 25 years of age: Thomas et al., 2007). Decline in performance was observed in older adulthood, occurring irrespective of stimulus type (e.g., prosodic speech; Mill, Alik, Realo, & Valk, 2009; nonverbal vocalizations; Lima et al., 2014; synthetic speech syllables; Bellis, Nicol, & Kraus, 2000). Whereas some studies support a positive bias in aging (i.e., relatively lower decline for positive vs. negative emotions; Mill et al., 2009), others report a similar decline in the recognition of negative and positive emotions (Lima et al., 2014; Paulmann, Pell, & Kotz, 2008; Ruffman, Halberstadt, & Murray, 2009).

Relative to the study of facial affect processing, fewer studies have examined how age affects emotion decoding in the auditory modality. Despite increasing evidence for a nonlinear development of emotion recognition, that is, accuracy improving and then declining at faster rates for some discrete emotions as compared to others (e.g., Grosbras et al., 2018; Lima et al., 2014), this evidence has been mixed, with some studies showing a decline in performance as a function of age (e.g., Dupuis & Pichora-Fuller, 2015; Lima et al., 2014). The examination of age-related differences in vocal emotion recognition across the life span is currently lacking. Additionally, most studies have relied on voice samples produced by adults (Chronaki et al., 2015; Grosbras et al., 2018), disregarding that emotion recognition might be influenced by the speaker’s age. The notion that age—the age of the person who expresses an emotion—affects emotion perception is partially supported by studies reporting age-related perceptual biases in visual processing, such as the own-age bias: participants are more accurate at recognizing own-age faces than other-age faces (children: Hills & Willis, 2016; young adults: Wiese, Schweinberger, & Hansen, 2008; older adults: Anastasi & Rhodes, 2005). Two meta-analyses have supported an own-age bias in children and young adults (children: 8 studies; young adults: 25 studies; Rhodes & Anastasi, 2012) as well as in older adults (16 studies; Martschuk & Sporer, 2018). Nonetheless, some studies failed to observe similar effects (children: Griffiths, Penton-Voak, Jarrold, & Munafò, 2015; adolescents: Vetter, Drauschke, Thieme, & Altgassen, 2018; young adults: Ebner & Johnson, 2009; older adults: Denkinger & Kinn, 2018).

Although studies probing an own-age bias in facial emotion recognition are both scarce and inconsistent, it is worth noting that this sort of bias might reflect a preference for in-group versus out-group features (Anastasi & Rhodes, 2005), which have been associated with enhanced emotion recognition (Elfenbein & Ambady, 2002). Thereby, an own-age bias in recognition could reflect familiarity effects: individuals may interact more with peers from the same age group as it is more likely that they share the same interests. An increased number of interactions with own-age peers might result in facilitated decoding of perceptual patterns adopted by that group to communicate emotions (Rhodes & Anastasi, 2012; Sen, Isaacowitz, & Schirmer, 2018). Accordingly, evidence for an own-age bias in emotion processing was observed in adolescents (Easter et al., 2005; Picci & Scherf, 2016), young adults (Ebner, He, & Johnson, 2011; Ebner et al., 2013; Riediger, Voelkle, Ebner, & Lindenberger, 2011), and older adults (Riediger et al., 2011). It is plausible that a close age match between the participant and the poser (maximum 1–3 years difference—Hills, 2012) is required for an own-age bias to be observed, which could explain the negative findings in some studies (Ebner & Johnson, 2009).

The few studies that probed how the speaker’s and listener’s ages interact during vocal emotion recognition also did not report consistent findings (Dupuis & Pichora-Fuller, 2015; Morningstar, Ly, Feldman, & Dirks, 2018; Sen et al., 2018; Table 1). Whereas some failed to document an own-age bias in recognition accuracy (Dupuis & Pichora-Fuller, 2015; Morningstar, Ly, et al., 2018), others provided evidence for a young-age bias. For example, both younger and older adult listeners were more accurate at decoding emotions produced by younger speakers (Sen et al., 2018). This bias could reflect an additive effect of familiarity and social motivation (Sen et al., 2018): young and older speakers rely on different acoustic parameters to convey emotion. Young listeners might be more familiar with the acoustic patterns used by young (in-group) speakers. Older listeners are familiar with the acoustic patterns used by younger speakers as well, due to accumulated experience with younger speakers.

### Table 1

<table>
<thead>
<tr>
<th>Study</th>
<th>Emotional categories</th>
<th>Listener age (M ± SD)</th>
<th>Speaker age (M ± SD)</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Younger</td>
<td>Older</td>
<td>Younger</td>
</tr>
<tr>
<td>Dupuis and Pichora-Fuller (2015)*</td>
<td>Anger; sadness; fear; neutral</td>
<td>19.7 ± 2.7</td>
<td>69.8 ± 4.5</td>
<td>26</td>
</tr>
<tr>
<td>Morningstar, Ly, Feldman, and Dirks (2018)*</td>
<td>Anger; disgust; fear; happiness; sadness*</td>
<td>21.6 ± 3.3</td>
<td>70.7 ± 4.0</td>
<td>26</td>
</tr>
<tr>
<td>Sen, Isaacowitz, and Schirmer (2018)</td>
<td>Anger; happiness; neutral; sadness</td>
<td>14.08 ± 0.75</td>
<td>21.24 ± 2.40</td>
<td>Female: 12.97 Male: 12.51</td>
</tr>
</tbody>
</table>

*SD age information missing. * The stimulus set also contained stimuli portraying friendliness and meanness, but these categories were not included in the table as they do not represent emotion categories.
social interactions over the years. Additionally, they may also consider interactions with stimuli produced by younger speakers as more rewarding (Gordon & Arvey, 2004; Sen et al., 2018).

The validity of an own-age bias in emotion recognition remains an open question, particularly in the auditory domain. Of note, no studies have examined such potential effects using nonverbal emotional vocalizations. Vocalizations such as laughter, cries, or screams are more primitive than speech prosody and are considered to be the auditory equivalent of facial expressions, that is, an “auditory face” (Belin, Fecteau, & Bédard, 2004). Vocalizations encode affective meaning without potential confounds associated with concurrent verbal content such as in speech prosody (Belin et al., 2004).

The study of nonverbal vocalizations is particularly relevant in light of evidence supporting developmental differences in emotion recognition as a function of stimulus type (Morningstar, Nelson et al., 2018; Sauter, Panattoni, & Happé, 2013). For example, whereas children were found to reach adult-level performance in emotion decoding from speech prosody by 8.5 years (Doherty-Fitzsimons, Asenbauer, & Staunton, 1999), adult-level performance was only reached at 15 years in the case of nonverbal vocalizations (Grosbras et al., 2018). Moreover, there is additional evidence that nonverbal vocalizations are processed differently from other auditory stimuli: Sauter and colleagues (2013) showed that children are better at decoding emotions from nonverbal vocalizations than from speech from an early age, whereas Pell and colleagues (Pell et al., 2015) also reported an advantage of nonverbal vocalizations versus speech prosody in adults.

The current study probed for the first time how accurately listeners in different stages of development (children, adolescents, young adults, older adults) recognize nonverbal vocal emotions and whether this capacity is enhanced when speakers and listeners belong to the same age group. Not only did the stimulus set include vocalizations produced by speakers of different age groups, but it also covered a wide range of positive and negative emotions (pleasure, relief, achievement, happiness, sadness, disgust, anger, fear, surprise).

The nature and magnitude of age-related differences in the recognition of vocal emotions, from childhood until older adulthood, were examined. Overall, these differences were expected to vary by emotion category, with some emotions showing accentuated improvement (from childhood to adolescence and young adulthood: e.g., anger and fear) or attenuated decline (from young adulthood to older adulthood: e.g., fear; Dupuis & Pichora-Fuller, 2015; Ruffman et al., 2008). We would support the notion that discrete emotions develop differently across the life span, a hypothesis grounded in previous studies employing a cross-sectional design to compare different age groups (Chronaki et al., 2015; Sauter et al., 2013). Nonetheless, our hypotheses were exploratory regarding the specific nature and magnitude of the expected age-related differences in vocal emotion recognition, considering the variability of findings from previous studies (e.g., Lima et al., 2014; Mill et al., 2009; Paulmann et al., 2008; Ruffman, Halberstadt, et al., 2009).

An own-age bias in vocal emotion recognition was also tested. We hypothesized enhanced emotion recognition accuracy for vocal emotions expressed by speakers of the same age group as the listeners. This hypothesis was grounded in evidence that partially supports an own-age bias in facial emotion recognition (Easter et al., 2005; Ebner et al., 2011; Riediger et al., 2011), which could arguably extend to nonverbal vocalizations as these have been thought to represent the auditory equivalent of facial expressions (Belin et al., 2004). It is also supported by prior auditory studies showing that exposure to specific patterns of sounds results in the development of different auditory profiles (Tervaniemi, JANUNEN, KRUCK, PUTKINEN, & Huotilainen, 2016). When individuals are more exposed to sounds produced by their own-age peers, they may show an auditory tuning to the acoustic patterns adopted by their peers. However, given previous evidence from emotional speech prosody studies showing no own-age biases (Dupuis & Pichora-Fuller, 2015; Morningstar, LY, et al., 2018), here we address this question in an exploratory manner.

Considering acoustic voice changes across the life span (e.g., Stathopoulos, Huber, & Sussman, 2011) and differences in the use of acoustic cues to decode emotions as a function of aging (e.g., Lima et al., 2014; Paulmann et al., 2008), we further tested whether accuracy in vocal emotion recognition as a function of age was predicted by acoustic cues of emotion (e.g., Justlin & Laukka, 2003). We hypothesized age-related differences in how the most salient acoustic cues (e.g., fundamental frequency [F0]) are used to decode emotions.

Method

Participants

A total of 164 participants (age range 7–82 years) took part in the study: 36 children (18 females; M_age = 8.7, SD = 1.1, age range 7–11 years), 53 adolescents (30 females; M_age = 15.0, SD = 1.4, age range 12–18), 48 young adults (24 females; M_age = 22.4, SD = 2.2, age range 20–30 years), and 27 older adults (17 females; M_age = 66.07, SD = 5.6, age range 58–82 years). The selection of age ranges is consistent with previous developmental/own-age bias studies (Chronaki et al., 2015; Grosbras et al., 2018; Hills & Willis, 2016; Lima et al., 2014). The age range defined for the children group was based on previous studies (e.g., Hills & Willis, 2016). The adolescent group only included participants in mid- or late-adolescence phases (e.g., Grosbras et al., 2018) to be more clearly separated from the children group. The age range of the older adult group is also similar to previous work (e.g., Lima et al., 2014). The older adults completed the Montreal Cognitive Assessment test, which ruled out the potential presence of cognitive decline (Freitas, Simões, Martins, Vilar, & Santana, 2010; Simões et al., 2008). All participants were native speakers of European Portuguese. Participants were selected by word of mouth and via research collaborations with schools and senior universities established by the hosts of the study (University of Minho and University of Lisbon). We did not conduct an a priori sample size calculation as this study was exploratory. Notwithstanding, we ensured that each age group included an approximate minimum of 30 participants following previous studies (e.g., Paulmann et al., 2008).

The study was conducted in accordance with the Declaration of Helsinki and was approved by the local ethics committee of the University of Minho (Braga, Portugal) and of the Faculty of Psychology-University of Lisbon (Lisbon, Portugal).
Stimuli

The stimuli were selected from a newly recorded corpus, developed specifically for this study. Short nonverbal vocal samples were produced by 10 children (8–11 years), 10 adolescents (14–16 years), 10 young adults (19–23 years), and 10 older adults (60–75 years), expressing nine different emotions (pleasure, relief, achievement, happiness, sadness, disgust, anger, fear, surprise), and a neutral state. All the recordings were made with the same recording equipment (Edirol R-09 recorder [Shure Incorporated, Cd. Juarez, Mexico] and Shure PG48 microphone [Roland Corporation, Shanghai, China]) in a quiet room. Speakers were instructed to produce brief vocalizations without any verbal content and to refrain from producing interjections (e.g., “Yeah,” “Yupi”). They were also told that they should try to produce the vocalizations in a natural and spontaneous way and to utter more than one vocalization per emotion. To facilitate the generation of vocal emotions and to ensure that all speakers interpreted the emotional labels in a similar manner, speakers were presented with both videos and scenarios related to each emotion and were asked to produce a nonverbal vocalization congruent with those videos/scenarios. For children speakers, age-appropriate videos and scenarios were presented to facilitate their performance and ensure that the emotional expression corresponded to the emotional label (see the online supplemental materials).

The recordings were edited to exclude, for example, nonemotional physiological sounds (i.e., sneezing) and background noise. The same editing steps were used for all recordings: first, background noise was removed using Version 2.1.1 of Audacity software (Audacity Team, 2014); intensity was normalized (i.e., 70 dB) using Praat software Version 5.1.05 (Boersma & Weenink, 2009). The final step was to individually segment the sounds using Praat (Boersma & Weenink, 2009). Three judges individually selected vocalizations from different speakers that they considered to best represent each emotional category. The selected vocal samples were those for which the judges came to a consensus. The final set consists of vocal samples from two male and two female speakers from each age group (16 in total). Mean duration (Table 2) and pitch (Table 3) were extracted using the R package soundsgn (Anikin, 2019). Pitch contours were checked manually against spectrograms. In addition, for each frame of the short-time Fourier transform we measured harmonics-to-noise ratio, spectral center of gravity (centroid), and subjectively perceived loudness in sone (assuming a Sound Pressure Level of 70 dB relative to a reference pressure of 20 μPa). These acoustic measures were then summarized per stimulus as the median and standard deviation of frame-by-frame values. Prosodic and voice quality characteristics such as these are highly salient to listeners and have previously been used to model emotion recognition in nonverbal vocalizations (e.g., Justlin & Laukka, 2003).

To avoid fatigue, the selected stimuli were divided into two lists with 80 vocalizations each. For each list, we ensured that, for a given emotion, there was an equal number of vocal samples produced by a female and by a male speaker of each age group (i.e., 8 vocalizations per emotion category in each list), presented in a randomized way. Due to some constraints (e.g., participant’s available time), some participants completed only one of the lists (trials = 80; see the online supplemental materials). No data were excluded from the analyses.

Procedure

A forced-choice emotion categorization task was used to examine how each of the nonverbal vocal sounds was assigned a specific emotional (pleasure, relief, achievement, happiness, sadness, disgust, anger, fear, surprise) or neutral label.

Before the experiment, participants were given instructions regarding the task, including an illustrative trial, and underwent a short training session. To ensure that participants from all age groups had the same understanding of each emotional label, the categorization task was adapted to the participant’s age. For example, besides the training phase, children were also presented with scenarios and vocal samples of each emotional label (selected from the Montreal Affective Voices battery; Belin, Filion-Bilodeau, & Gosselin, 2008; see the online supplemental materials). In the case of adolescents, each emotional label was explained to them by the experimenter, who provided specific examples clarifying the meaning of each label. Each trial started with the presentation of a fixation cross for 500 ms, followed by a vocalization. After listening to the vocalization, participants were given the choice to either listen to it again or to proceed to its categorization. There was no limit for the number of repetitions. After proceeding to the

Table 2
Mean Duration of Vocalizations (Milliseconds) per Emotion and Age Group

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Children (M ± SD)</th>
<th>Adolescents (M ± SD)</th>
<th>Young adults (M ± SD)</th>
<th>Older adults (M ± SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Achievement</td>
<td>1,011 ± 162</td>
<td>1,085 ± 297</td>
<td>1,693 ± 1,176</td>
<td>1,118 ± 324</td>
</tr>
<tr>
<td>Anger</td>
<td>1,351 ± 493</td>
<td>709 ± 94</td>
<td>742 ± 301</td>
<td>1,252 ± 359</td>
</tr>
<tr>
<td>Disgust</td>
<td>778 ± 280</td>
<td>800 ± 252</td>
<td>845 ± 159</td>
<td>916 ± 224</td>
</tr>
<tr>
<td>Fear</td>
<td>1,056 ± 610</td>
<td>555 ± 167</td>
<td>478 ± 363</td>
<td>512 ± 99</td>
</tr>
<tr>
<td>Happiness</td>
<td>1,483 ± 139</td>
<td>999 ± 263</td>
<td>2,461 ± 613</td>
<td>1,338 ± 522</td>
</tr>
<tr>
<td>Neutral</td>
<td>930 ± 166</td>
<td>910 ± 114</td>
<td>598 ± 301</td>
<td>853 ± 602</td>
</tr>
<tr>
<td>Pleasure</td>
<td>1,484 ± 169</td>
<td>1,362 ± 481</td>
<td>1,113 ± 285</td>
<td>1,592 ± 474</td>
</tr>
<tr>
<td>Relief</td>
<td>1,669 ± 567</td>
<td>775 ± 69</td>
<td>859 ± 598</td>
<td>1,376 ± 443</td>
</tr>
<tr>
<td>Sadness</td>
<td>1,099 ± 317</td>
<td>1,636 ± 785</td>
<td>3,081 ± 1,057</td>
<td>993 ± 377</td>
</tr>
<tr>
<td>Surprise</td>
<td>896 ± 187</td>
<td>819 ± 310</td>
<td>446 ± 20</td>
<td>1,017 ± 250</td>
</tr>
</tbody>
</table>

categorization, participants were presented with 10 possible options to categorize each vocalization (i.e., achievement, anger, disgust, happiness, fear, neutral, pleasure, relief, sadness, and surprise) and responded by pressing a keyboard key. Each response option was accompanied by a correspondent cartoon exemplifying the emotional label, which simplified the task for the younger participants and was suitable for all age groups. There was an interval of 1,000 ms between trials. Vocalizations were presented using Superlab software (Cedrus, San Pedro, CA).

**Data Analysis**

The statistical model adopted in the current study assumes that each participant has an idiosyncratic base level of recognition (independently of age), and that certain stimuli are harder to decode than others (independently of emotion). This means that the statistical model takes into consideration the existence of variability both within the participants and the stimuli. Hence, we did not remove any participants or stimuli from the statistical analyses.

All models used a single participant’s response to a single stimulus as the outcome, without aggregating the data. To account for nested structure, accuracy was modeled with multilevel logistic regression with various main effects and three random intercepts: per stimulus, speaker, and listener. The effect of the listener’s age on the accuracy of emotion recognition was nonlinear and not easily captured by quadratic models on the original or log-transformed scale. Therefore, we opted for spline smoothing with Bayesian generalized additive mixed models. All mixed models were fit in R with the brms package (Bürkner, 2017) using default, mildly normalizing priors. The results were summarized as the median of the posterior distribution and 95% credible intervals. Nested models were compared using the widely applicable information criterion (WAIC), which is a measure of a model’s predictive power penalized for complexity to avoid overfitting: models with a lower WAIC are preferred in the sense that they are expected to make better predictions of future data (Watanabe, 2010). Of note, the neutral category was treated in the same way in the analyses than the other emotional categories, following previous studies (Chronaki et al., 2015; Sen et al., 2018).

To predict the perceived emotion based on acoustic characteristics of the stimuli, we built random forest models (Breiman, 2001). Random forests are powerful classifiers based on decision trees that are particularly suitable for analyzing large data sets with multiple predictors and potential complex interactions. Classification accuracy was estimated using simple Monte Carlo cross-validation with 100 iterations: for each iteration, the dataset was split into equal-sized training and test sets, and the accuracy in the test set was averaged across all iterations.

**Results**

**Effects of the Listener’s Age**

As shown in Figure 1, recognition accuracy initially increases with the listener’s age, starting from 55.9% (95% CI [46.8, 64.5]) at the age of 7 and peaking at 72.8% (95% CI [66.1, 78.6]) at the age of 20. By the age of 60, accuracy declines to 48.0% (95% CI [39.1, 57.4]), which is lower than for 7-year-olds. After 60, the decline in recognition accuracy becomes even more pronounced, with the fastest acceleration (minimum second derivative) at the age of 67 and the fastest rate of deterioration (the most negative slope) at the age of 75. According to the available data on elderly participants, by 75 years the overall accuracy is predicted to be just 22.0% (95% CI [14.3, 30.9]), barely above the chance level of 10%.

The overall accuracy across all ages did not depend on the listener’s sex: women were, on average, only 0.9% (95% CI [−4.0, 5.7]) more accurate than men. In addition to potential sex differences, the age-related changes in recognition accuracy may be emotion specific. To compare the effect of listener age on the recognition accuracy for different emotions, we fit a model predicting recognition accuracy in individual trials with a separate smooth term for each emotion. The model probing age-related emotion-specific differences was strongly preferred based on WAIC (19,788.5 without emotion and 19,501.6 with emotion, difference = 286.9, SE = 37.9). As shown in Figure 2, for all emotions there was a tendency for recognition accuracy to increase with the listener’s age up to a certain point, followed by a gradual decline. However, the magnitude of age-related changes in recognition accuracy varied per emotion. An improvement from childhood (Age 9) to adolescence (Age 15) was found both overall (10.1%, 95% CI [6.0, 14.1]) and for most individual emotions, with the exception of pleasure (2.9%, 95% CI [−1.0, 7.5]), relief (3.8%, 95% CI [−0.5, 9.3]), sadness (2.4%, 95% CI [−3.0, 8.4]), and surprise (0.5%, 95% CI [−2.6, 5.7]), for which recognition accuracy was relatively stable during this developmental period (see Figure 2). Little further improvement was seen from adolescence (Age 15) to young adulthood (Age 23) overall (2.4%, 95% CI [−1.5, 6.2]), although accuracy did increase significantly for some emotions, such as fear (8.9%, 95% CI [4.3, 14.1]) and achievement (6.6%, 95% CI [1.9, 11.2]). Comparing the ages of 23 versus 60, there was a marked overall drop in performance (−23.4%, 95% CI [−30.5, −16.8]), which was also observed for all individual emotions except fear (2.8%, 95% CI [−5.9, 11.5]) and surprise (−6.4%, 95% CI [−14.1, 0.7]). In contrast, the decline in accuracy from 23 to 60 was particularly dramatic for pleasure (−46.1%, 95% CI [−54.1, −36.6]) and sadness (−30.8%,
95% CI [−40.3, −20.6]). Comparing the ages of 60 versus 75, there was a further significant deterioration in the recognition of emotion both overall (−25.4%, 95% CI [−35.4, −15.5]) and for most individual emotions, except sadness (2.4%, 95% CI [−13.4, 22.0]) and surprise (−2.0%, 95% CI [−13.8, 12.2]; Figure 3). Notably, the decline of recognition accuracy from 60 to 75 years of age was of similar magnitudes for positive emotions (19.2%, 95% CI [11.7, 27.3]; averaging across achievement, happiness, pleasure, and relief) and negative emotions (20.1%, 95% CI [10.2, 29.6]; averaging across anger, disgust, fear, and sadness).

Effects of the Speaker’s Age

The fitted overall recognition accuracy for each of four speaker age groups, averaging across listeners of all ages, was as follows: children: 71.8% (95% CI [60.2, 80.2]); adolescents: 62.1% (95%
The current study reports, for the first time, changes in vocal emotion recognition from childhood to older adulthood using nonverbal vocalizations. Furthermore, it probed whether and how the speaker’s and listener’s ages interact when decoding the emotional voice quality, and specifically whether these interactions reflect an own-age bias. Listeners of different ages (children, adolescents, young adults, older adults) were instructed to explicitly categorize vocal emotions produced by speakers of the same or of a different age group in a forced-choice task.

Our findings reveal that emotion recognition accuracy across the life span is better characterized by a nonlinear function (Grosbras et al., 2018), with a large improvement from childhood to adolescence (Chronaki et al., 2015, 2018; Grosbras et al., 2018), and a decline in older adulthood (Dupuis & Pichora-Fuller, 2015; Lima et al., 2014; Ruffman et al., 2008). The degree of improvement versus decline differed by emotion, supporting the existence of emotion-specific age-related differences (Chronaki et al., 2015; Isaacowitz et al., 2007). Recognition accuracy was further modulated by the speaker’s age: the worst performance was observed when emotions were produced by older adults, replicating previous studies (Dupuis & Pichora-Fuller, 2015). Furthermore, only children’s performance accuracy increased when vocal emotions were produced by speakers of the same age group (own-age bias). No effects involving either listener or speaker sex were found, that is, men and women were similarly good at recognizing emotions, regardless of whether they were produced by male or female speakers (Chronaki et al., 2015; Sauter et al., 2013). Of note, we tried to ensure that all participants understood each emotional label in a similar manner to minimize potential effects of differences in verbal ability on categorization performance.

Emotion recognition accuracy increased from childhood to adolescence for most emotions (Chronaki et al., 2015; Grosbras et al., 2018), except for pleasure, relief, sadness, and surprise. This confirms that the recognition of vocal emotions is an early developing mechanism (Chronaki et al., 2018), particularly for the latter categories. The faster increment in sadness recognition agrees with prior studies (Grosbras et al., 2018; Vidas, Dingle, & Nelson, 2018). It is also consistent with studies showing that infants as young as 3 to 7 months old show an enhanced brain response to vocalizations expressing sadness but not happiness. Specifically, increased activation was observed in the orbitofrontal cortex and the insula in response to sad but not happy (vs. neutral) nonverbal vocalizations (Blasi et al., 2011). Of note, these brain regions are consistently engaged in emotional processing in adults (Blasi et al., 2011; Leppänen & Nelson, 2009). Moreover, the faster development of sadness recognition may be due to the relevant role that sadness plays in social behavior learning (Blasi et al., 2015). Vocal emotion recognition continues to develop (although less prominently) beyond childhood, from adolescence to adulthood. Brain maturation and socialization factors may contribute to changes in vocal emotion decoding observed in adolescence (Blakemore, 2008). Previous studies with nonverbal vocalizations reported an adult-level performance by Age 15 (Grosbras et al., 2018; M_{age} of the adolescent group in the current study = 15.0 years). However, considerable improvement was observed for fear and achievement.

Discussion

EMOTION RECOGNITION ACROSS THE LIFE SPAN

Using the top five acoustic predictors of perceived emotion (stimulus duration, mean pitch [F0] and harmonicity, and the variability in spectral centroid and loudness), a random forest model correctly predicted 62.1% of participants’ responses. This is slightly higher than the accuracy with which the listeners themselves recognized the portrayed emotion (59.8%), suggesting that we succeeded in capturing the most salient acoustic characteristics of the stimuli. Including the listener’s age group as a predictor did not noticeably improve the prediction accuracy (from 62.1% to 62.4%). Random forest models are highly efficient at uncovering complex interactions between multiple predictor variables. If there were important age-specific differences in the utilization of acoustic cues of emotion, we would expect models with the listener’s age to be considerably more accurate at predicting the perceived emotion. No such improvement was observed, indicating that acoustic cues of emotion were used in a broadly consistent manner by listeners of all ages.

Acoustic Predictors

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
Kleef, Fischer, & van der Schalk, 2009). As the same number of
and disgust) are more easily decoded than others (Hawk, van
These differences indicate that certain emotions (e.g., happiness
older adulthood, recognition accuracy varied by emotion type.
We also believe that auditory deficits do not explain the current
results as normative decline in auditory measures was found not to
Whereas we observed improved emotion recognition from
childhood to adulthood, and a decline from young adulthood to
older adulthood, recognition accuracy varied by emotion type. These differences indicate that certain emotions (e.g., happiness and disgust) are more easily decoded than others (Hawk, van Kleef, Fischer, & van der Schalk, 2009). As the same number of positive and negative emotions was included in the current experiment, the improved recognition of happiness cannot be explained by the use of valence information during categorization (Grosbras et al., 2018). Instead, the higher recognition accuracy may be related to the high social relevance of laughter (representing happy vocalizations in the current study): laughter plays a critical role in establishing and maintaining social bonds, and in the management of negative experiences (Scott, Lavan, Chen, & McGettigan, 2014). Additionally, the high recognition of disgust may be accounted for by its survival value: vocalizations expressing disgust may signal a threat for the organism such as rotten food (Fessler & Haley, 2006; Zimmer, Höfler, Koschutnig, & Ischebeck, 2016). Most importantly, the absence of improved recognition for happiness and disgust from adolescence to adulthood (Grosbras et al., 2018; Morningstar, Ly, et al., 2018) as well as the general decline observed between young and older adulthood (Lima et al., 2014) are consistent with previous studies.

Contrary to our hypothesis, age-related differences in performance were not accounted for by differences in the use of acoustic cues to decode emotion. The analysis using random forest showed that listeners of different ages interpreted the most salient acoustic cues of emotion in a generally similar manner. This suggests similar acoustically based inference rules across age. Instead, the overall decline in emotion recognition in older adults could be accounted for by age-related changes in brain regions that are engaged in the processing of emotional cues (Ruffman et al., 2008). For example, age-related volume decreases were observed in the frontal cortex, but seem to be strongest in the orbitofrontal cortex (Lamar & Resnick, 2004; Tisserand et al., 2002), which is implicated in evaluative judgments of vocal emotions (Schirmer & Kotz, 2006). These changes could partially explain deficits in the cognitive evaluation of the emotional significance of the voice as a function of aging (Ruffman et al., 2008), particularly of anger (Sander et al., 2005). Of note, the delayed decline for fear and surprise is consistent with previous studies, and it may reflect a processing advantage due to the motivational relevance of these emotions: both fear and surprise vocalizations might signal possible threats, thus capturing attention more easily (Dupuis & Pichora-Fuller, 2015).

Considering the speaker’s age effects, we observed that listeners were less accurate at decoding vocal emotions produced by older adults (for similar results, see Dupuis & Pichora-Fuller, 2015), whereas they were more accurate at decoding emotions produced by children and young adults. An interaction between speaker and listener ages revealed a significant positive own-age bias for children, a negative own-age bias for older adults (marginal effect), and no own-age bias for adolescents and young adults. Emotion recognition might be more flexible in childhood as the neurocognitive system is still undergoing major developmental changes, whereas it is more stable in adulthood and more reliant on automatic processes (Lima, Ankin, Monteiro, Scott, & Castro, 2019). A more flexible system could potentially be more permeable to familiarity effects (i.e., greater and recent exposure to own-age peers); while children may interact with individuals from all ages due to societal rules (e.g., enrollment in primary education), they may spend a greater amount of time interacting with peers of the same age and consequently be more exposed to own-age vocal stimuli. The greater exposure to own-age voices may result in specific changes in auditory processes and, conse-
quentiy, in processing advantages for this type of stimulus. The idea that auditory experience may guide perceptual biases (e.g., own age) is supported by studies probing the impact of musical expertise on auditory processing (Krishnan et al., 2018; Tervaniemi et al., 2016). Specifically, these studies show that musicians may develop different auditory profiles according to the musical genre (e.g., classical, jazz, or rock) or the instrument (e.g., guitar, vocal chords) they specialize in: they seem to be more sensitive to sounds that violate the musical template associated with their specific musical training (Tervaniemi et al., 2016) and respond differentially to sounds produced by the musical instrument they can play (Krishnan et al., 2018). Auditory experience could explain the bias observed in the current study: children might be more tuned to acoustic features of children’s voices, which could account for the facilitated decoding of vocal emotions produced by speakers of the same age. While we tried to ensure that all speakers and listeners interpreted the emotional labels in an identical manner, it is possible that greater exposure to own-age peers may result in similar interpretations of the emotional labels within an age-group, which could boost emotion recognition.

As previous studies probing speaker age effects on emotion recognition with prosodic speech did not include children either as listeners or speakers (Dupuis & Pichora-Fuller, 2015; Morningstar, Ly, et al., 2018; Sen et al., 2018), we cannot determine to what extent the observed own-age bias for children is specific to non-verbal vocalizations only. Examining how vocal emotions are decoded as a function of age has critical implications for our understanding of disorders characterized by altered emotional processing such as schizophrenia (Pinheiro et al., 2013). The resulting knowledge may also aid the development of interventions targeting impaired emotion recognition skills.

While the current study sheds light on the impact of listener and speaker ages on emotion recognition, some limitations should be noted. For example, there was an effort to minimize the potential effects of verbal ability differences between children and adolescents, which could impact their understanding of the emotional labels (e.g., Baron-Cohen, Golan, Wheelwright, Granader, & Hill, 2010). Notwithstanding, future studies should more directly control for potential effects of verbal ability on emotion categorization. Further, participants were not asked to report how frequently they interact with own- versus other-age persons, as well as their interest in interacting with persons from different age groups. Future studies should collect such additional measures to specify how familiarity contributes to the own-age bias. Along the same lines, considering that own-age bias might be dependent on a close match between speaker and listener ages (Hills, 2012), it is also possible that the absence of an own-age bias in adolescence and young adulthood reflects a mismatch between listener and speaker ages in the current study. Considering that surprise was not explicitly labeled as either “positive” or “negative,” future studies should clarify whether vocal recognition of surprise differs when its valence is specified. In the current experiment, participants were able to repeat the vocalizations as many times as they wanted, but the number of repetitions was not recorded. Future studies adopting a similar procedure should examine the effects of the number of stimulus repetitions on participant’s performance. Lastly, while we were able to recruit a high number of listeners, the sample of speakers used in this study was relatively small ($N = 16$). In order to enable higher precision regarding the estimation of an own-age bias, as well as changes in recognition during adulthood, future studies should consider using a larger sample of speakers.

**Conclusion**

The current findings provide further support for emotion-specific life span changes in vocal emotion recognition: they confirm a nonlinear pattern of improvement followed by decline in emotion recognition as a function of age. In addition, they show that speaker age may interact with listener age during emotion recognition. Specifically, children seem to be more attuned to and more accurate at decoding vocal emotions produced by speakers of the same age group (but such an own-age bias is absent in adolescence and adulthood). Together, these findings indicate that voice perception is not only a matter of how one speaks but also of who speaks and who listens. Age effects should be considered by current models of vocal emotional perception (e.g., Schirmer & Kotz, 2006).

**References**


Received April 23, 2019
Revision received September 4, 2019
Accepted September 4, 2019