J Nonverbal Behav (2017) 41:155–183
DOI 10.1007/s10919-017-0250-7

ORIGINAL PAPER

Vocal Cues Underlying Youth and Adult Portrayals of Socio-emotional Expressions

Michele Morningstar[1] · **Melanie A. Dirks**[1] · **Shiming Huang**[1]

Published online: 16 January 2017

Abstract Despite known differences in the acoustic properties of children's and adults' voices, no work to date has examined the vocal cues associated with emotional prosody in youth. The current study investigated whether child ($n = 24$, 17 female, aged 9–15) and adult ($n = 30$, 15 female, aged 18–63) actors differed in the vocal cues underlying their portrayals of basic emotions (anger, disgust, fear, happiness, sadness) and social expressions (meanness, friendliness). We also compared the acoustic characteristics of meanness and friendliness to comparable basic emotions. The pattern of distinctions between expressions varied as a function of age for voice quality and mean pitch. Specifically, adults' portrayals of the various expressions were more distinct in mean pitch than children's, whereas children's representations differed more in voice quality than adults'. Given the importance of pitch variables for the interpretation of a speaker's intended emotion, expressions generated by adults may thus be easier for listeners to decode than those of children. Moreover, the vocal cues associated with the social expressions of meanness and friendliness were distinct from those of basic emotions like anger and happiness respectively. Overall, our findings highlight marked differences in the ways in which adults and children convey socio-emotional expressions vocally, and expand our understanding of the communication of paralanguage in social contexts. Implications for the literature on emotion recognition are discussed.

Keywords Vocal cues · Emotional expression · Youth · Vocal communication

✉ Michele Morningstar
michele.morningstar@mail.mcgill.ca

[1] Department of Psychology, McGill University, 1205 avenue Dr. Penfield, Montreal, QC H3A 1B1, Canada

Introduction

Understanding social and emotional cues conveyed by others is a key component of successful interpersonal interactions. One important source of such information is the human voice. To date, most research on vocal communication has focused on the acoustic properties associated with adult speakers' portrayals of basic emotions, such as anger and happiness. Little is known about youth's communication of emotional states. It has been documented that adult and child voices differ in acoustic properties such as pitch levels and harmonics (Stathopoulos et al. 2011; Whiteside and Hodgson 2000), yet no work to date has examined the vocal cues associated with emotional prosody in children. Moreover, other types of social information can also be conveyed vocally, including signals of social affiliation or rejection. These expressions may play a crucial role in the social interactions of both youth and adults, but little work has examined the auditory cues that characterize them. In this study, we sought to address these questions by examining differences in adults' and children's vocal portrayals of key socio-emotional expressions, including basic emotions such as anger, happiness, and sadness, as well as portrayals of "meanness" (i.e., social rejection) and "friendliness" (i.e., social affiliation).

Crick and Dodge's (1994) social information processing model suggests that accurate interpretation of social cues is necessary for the enactment of an appropriate behavioral response. Crucially, the ability to recognize others' emotional states has been linked to positive social outcomes; for instance, children's ability to recognize nonverbal cues of emotion has been associated with greater social competence (Izard et al. 2001; Leppänen and Hietanen 2001; Maxim and Nowicki 2003; Trentacosta and Fine 2010). Information about others' emotional experience can be conveyed via various modalities, including facial expressions (Ekman et al. 1980; Keltner and Ekman 2000) and inflections of the voice. Speakers' voices provide reliable information about their emotional state (Banse and Scherer 1996; Bänziger et al. 2015; Juslin and Laukka 2003), even beyond the content of their speech (Scherer et al. 1991). A large body of research has identified patterns of vocal cues linked to specific emotions (Johnstone and Scherer 2000; Scherer 2003). For instance, happiness is typically expressed with a high pitch, high intensity, and a fast speech rate (Johnstone and Scherer 2000; Juslin and Laukka 2001, 2003), whereas sadness is characterized by low pitch, low intensity, and a slow speech rate (Johnstone and Scherer 2000; Juslin and Laukka 2003; Pittam and Scherer 1993; Scherer 1996).

These vocal cues play a crucial role in interpersonal communication from a young age. Children and adolescents are proficient at detecting emotional information from affective prosody (Eskritt and Lee 2003; Sauter et al. 2012) and, by 10 years of age, rely on vocal characteristics more than verbal content to identify speakers' intent (Friend 2000). Despite the importance of vocal cues in youth's understanding and interpretation of social interactions, virtually no work has examined the acoustic correlates of socio-emotional expressions in children's speech, and studies of children's recognition of vocal displays of emotion have often relied exclusively on adult-generated stimuli (e.g. Allgood and Heaton 2015; Brosgole and Weisman 1995; Chronaki et al. 2014; Friend 2000; Sauter et al. 2012; Zupan 2015). Children have different respiratory and laryngeal systems than adults (Fitch and Giedd 1999; Stathopoulos and Sapienza 1993); thus, their voices differ from adults' along important dimensions like pitch levels and harmonic properties (Stathopoulos et al. 2011; Sussman and Sapienza 1994; Whiteside and Hodgson 2000). Given these differences, it is likely that the emotional prosody of children also varies from that of adults, a possibility not yet tested explicitly. To understand youth's communication of emotional information accurately, it is critical to typify the acoustic correlates of children's emotional speech and determine whether they differ from adults'.

Our understanding of youth and adults' socio-emotional prosody would also be deepened by examining the vocal cues underlying an expanded set of expressions. Thus far, researchers have focused primarily on mapping the acoustic correlates of "basic" emotions, such as anger and happiness. However, the voice can convey important social information about attitudes or intentions towards others, such as sarcasm (Cheang and Pell 2008; Rockwell 2000), boredom (Scherer and Oshinsky 1977), and tenderness (Juslin and Laukka 2003). The accurate recognition of social expressions, including cues of social rejection and affiliation, is necessary for interpersonal success. Exposure to social rejection is deemed hurtful by youth (Paquette and Underwood 1999) and is linked to adjustment problems (Rigby 2000; Underwood 2003). Moreover, enhanced vigilance to cues of affiliation and rejection may lead to greater rejection sensitivity (Masten et al. 2009). Thus, cues to expressions of social rejection (or "meanness") or affiliation (or "friendliness") are integral aspects of social interactions for adolescents. Previous research has investigated similar interpersonal cues. For example, behavioral warmth was found to be conveyed most by smiles and positive statements made about others (Bayes 1972), and the behavioral correlates of social aggression included mean glares and faces, and turning away to exclude someone from a group (LaFrance 2002; Simmons 2002; Underwood 2004). These investigations did not include characteristics of the speakers' voices; thus, we sought to examine the vocal prosody associated with the expression of "meanness" and "friendliness" in the current study.

Goals and Hypotheses of the Current Research

The goal of this study was to compare children and adults' vocalization of the basic emotions of anger, disgust, fear, happiness, and sadness (Banse and Scherer 1996), as well as the social expressions of meanness and friendliness. We followed the "standard content paradigm", which is widely used in the study of the communication of socio-emotional expressions (see Juslin and Laukka2003). We recruited child and adult actors to portray the aforementioned expressions. The use of actors has been the preferred method for obtaining prototypical portrayals of emotions (Scherer 2003); compared to naturalistic capture of vocal expressions, this approach allows for greater control over the quality of recordings obtained and the content of speech. We created audio recordings of actors' performances and performed speech analysis to extract acoustic information about pitch (mean and range), intensity (mean and range), temporal information (speech rate), and a measure of voice quality (mean harmonics-to-noise ratio). This combination of vocal cues is commonly investigated in research on vocal paralanguage (e.g., Cheang and Pell 2008; Laukka et al. 2008), as pitch, intensity, and speech rate have been identified as important cues for listener identification of affective information in the voice (Bänziger et al. 2015; Juslin and Laukka 2003; Pittam and Scherer 1993; Scherer et al. 1972; Scherer and Oshinsky 1977). We also considered voice quality, as it has been implicated in emotion recognition (Gobl and Ní Chasaide 2000; Scherer and Oshinsky 1977; Zhang 2008). We chose to highlight range values of pitch and intensity rather than variation, because range is a voice dynamic parameter that is more noticeable to the human ear (Rochman et al. 2008). Thus, we opted to focus on cues that had precedent in the literature, were identified as crucial for the communication of emotional paralanguage, and were perceptually relevant to listeners.

Given known differences in adult and child vocal tract length (Fitch and Giedd 1999), we expected several age-related variations in vocal cues, such that children's voices would be higher in mean pitch than adults'. As well, consistent with research on adults' voices (Juslin and Laukka 2003), we expected that the expressions would differ from one another in vocal characteristics; for example, anger typically differs from sadness in pitch, intensity, and temporal measures (Johnstone and Scherer 2000). Our central hypothesis was that the pattern of vocal cues across expressions may vary as a function of age. Various aspects of emotional processing improve with age, such as emotion recognition (Chronaki et al. 2014), perceived control over one's emotions (Gross et al. 1997), and the ability to understand complex components of emotions (Pons and Harris 2005). Because of this, it is likely that the ability to convey emotional expressions would also develop over time. This ability may be manifested as greater differentiation between expressions, which would enable listeners to better decode one's emotional output and respond appropriately. Due to adults' greater experience communicating emotional cues over their lifespan, we expected that they would produce more differentiated and distinct portrayals of various expressions than children. Specifically, we hypothesized that these distinctions may be most evident for pitch variables, given pitch's crucial role in listeners' ability to distinguish emotions in recognition tasks (Pell et al. 2009; Scherer 1996).

Further, we investigated the communication of meanness and friendliness by comparing them to the acoustic representations of similar basic emotions. One might expect meanness to resemble anger, which is the emotion typically linked to aggression (e.g., Fives et al. 2011; Wilkowski and Robinson 2008); however, this association is imperfect and complex (Averill 1983; Hortensius et al. 2012; Wilkowski and Robinson 2008). Researchers distinguish between reactive aggression (borne of anger or frustration; Hubbard et al. 2002) and proactive aggression, which can be used to achieve instrumental goals (Dodge and Coie 1987). Given that interpersonal hostility is not always associated with anger, we expected that the acoustic cues underlying the expression of anger and meanness would also be different. Because meanness may derive from lesser emotional arousal than reactive anger, we hypothesized that meanness would be expressed with lower values of mean pitch and intensity than anger.

Similarly, though positive mood may facilitate a social-approach motivation (Cunningham1988), one could communicate affiliation, or be friendly, without experiencing happiness; as such, the acoustic characteristics of both expressions may be different. To date, only one investigation has examined the acoustic correlates of "friendly-sounding speech" (Noble and Xu 2011), testing how vocal cues predicted listeners' ratings of how "friendly" and "happy" speech recordings sounded. Pitch mean and range increases were associated with higher ratings of both friendliness and happiness, but the increases were more noticeable for happiness. For this reason, we expected friendliness to be expressed with lower pitch mean and range than happiness.

Method

Participants

Child Actors

We recruited 24 English-language dominant child actors (17 females, 7 males), aged 9–15 years old ($M = 12.84$ years, $SD = 1.69$), from theatre schools and acting agencies in a large Canadian city. Child actors had 1–7 years of acting experience ($M = 3.25$ years,

$SD = 1.62$). The Pubertal Development Status self-report questionnaire was used to assess their developmental status (Petersen et al. 1988). Of the female actors, 1 self-reported as ‘prepubertal’, 1 as ‘beginning pubertal’, 5 as ‘midpubertal’, 8 as ‘advanced pubertal’, and 2 as ‘postpubertal’. Of the male actors, 3 identified as ‘beginning pubertal’, 2 as ‘midpubertal’, and 2 as ‘advanced pubertal’.

Adult Actors

We recruited 30 English-language dominant adult actors (15 females, 15 males), aged 18–63 years old ($M = 29.30$ years, $SD = 14.14$), from the same city. Actors had 1.5–35 years of acting experience ($M = 9.77$ years, $SD = 8.00$).

Procedure

All procedures were approved by the local Research Ethics Board. Adult actors gave written consent; written parental consent and written assent were obtained for child actors. Following standard content paradigm procedures (Juslin and Laukka 2003), each actor portrayed the same five sentences in the following expression conditions: anger, disgust, fear, friendliness, happiness, meanness, sadness, and neutral (see Fig. 1 for summary of experimental protocol). The sentences were constructed to be neutral in meaning, social in nature, and applicable to each of the expressions. Though researchers often ask actors to perform one-word utterances or neutral content sentences, such as “It is 11 o’clock” (e.g., Hammerschmidt and Jürgens 2007; Leinonen et al. 1997), we included sentences that would occur in socially significant interactions (e.g., “Why did you do that?”) with the goal of capturing typical examples of emotional speech in a social context.

Actors were fitted with a head-mounted microphone, held at a constant distance of approximately 5 cm with a 45° angle from the speaker’s mouth (Scherer et al. 1991). We first recorded actors’ performance of each sentence in a neutral tone of voice, defined as “your everyday voice, with no emotion” (c.f. Cheang and Pell 2008; Juslin and Laukka 2001). Next, actors performed the sentences in each of the seven expression conditions, aided by situational vignettes depicting social interactions in which the protagonist said the required sentence to a peer. The use of vignettes in production studies is common (e.g., Banse and Scherer 1996; Scherer et al. 1991) and facilitates actors’ task by providing a context for the expression they are asked to produce. To avoid the effects of reading on vocal prosody (e.g., Blaauw 1994; Howell and Kadi-Hanifi 1991), actors memorized the sentence to be performed, which was then hidden from their view. To facilitate the task, stimuli presentation was blocked by expression, such that all sentences were performed within each expression before moving to the next condition. The order of expression blocks, as well as the order of sentences within each block, was pre-randomized. We created two takes of each sentence in each condition, and participants who wished to create additional takes were permitted to do so. Actors were then debriefed and compensated for their time.

The above procedure resulted in 4353 recordings (54 participants × 8 expressions × 5 sentences × 2 takes, with 33 additional takes made by participants). Due to their small number, we removed the 33 additional takes from analysis. As well, 90 recordings were missing due to problems in the quality of the recording ($n = 56$), the actors saying the wrong words ($n = 30$), and experimenter errors ($n = 4$).

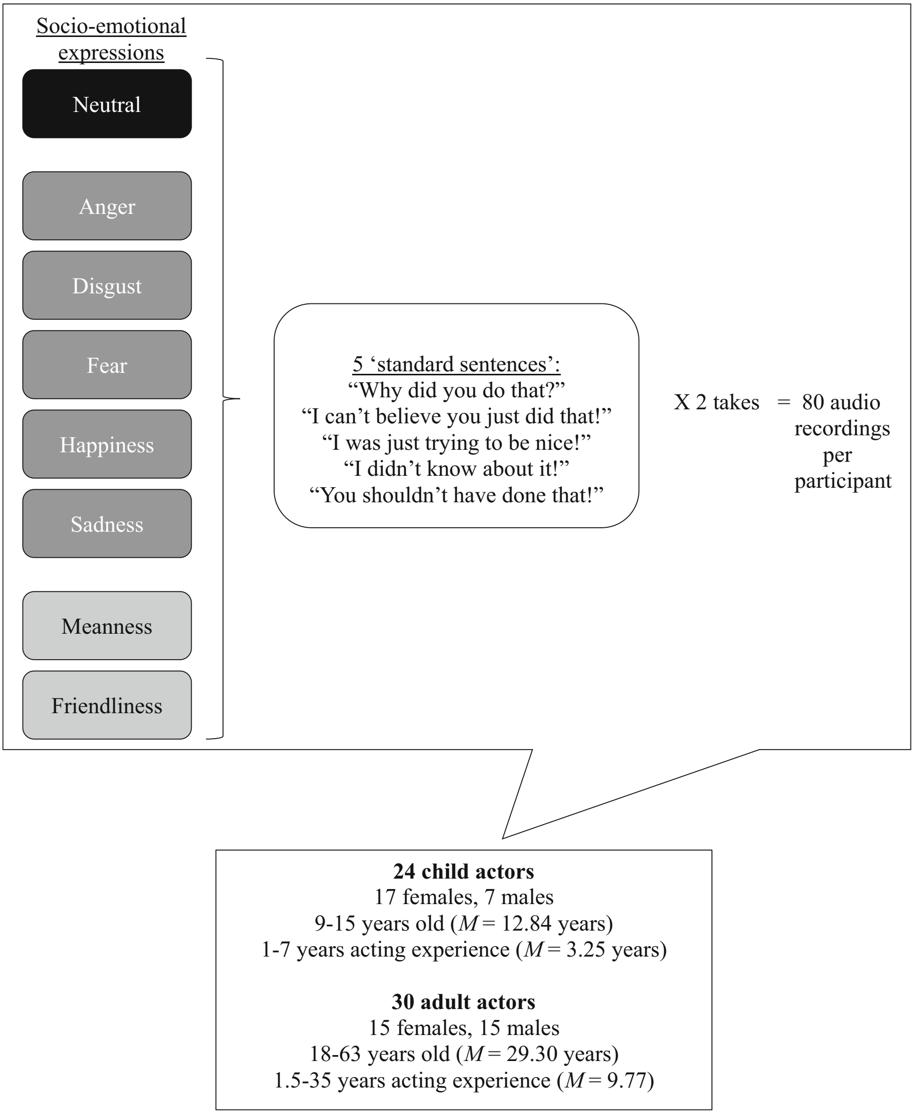

Fig. 1 Experimental protocol for recording child and adult actors' performance of socio-emotional expressions

Validity of the Recordings

As part of the standard content procedure (Johnstone and Scherer 2000; Juslin and Laukka 2003), we wished to ensure that these recordings accurately represented the intended expressions. Thus, we recruited 86 adult listeners (51 females, 35 males), aged 18 to 30 years old ($M = 21.24$ years, $SD = 2.40$), and 61 child listeners (34 females, 27 males), aged 9 to 16 years old ($M = 13.59$ years, $SD = 1.45$), to determine whether their ability to identify the intended expression in actors' recordings was consistent with accuracy levels typically reported in the emotion recognition literature.

The complete list of recordings produced by the actors ($n = 3807$) was narrowed to a more manageable sample for validation based on ratings of authenticity and recognizability provided by 6 independent raters (50% female; aged 20–25, $M = 22.33$ years, $SD = 2.16$), following a procedure similar to that of Banse and Scherer (1996). All listeners and raters spoke English as their dominant language and had no hearing or speech communication deficits. The 140 recordings with the highest ratings on both dimensions for each expression (except neutral), group of actors (men, women, boys, girls), and sentence were retained for inclusion in the sample provided to listeners.

Listeners heard all 140 recordings twice, in a randomized order, over sound-cancelling headphones. They were asked to indicate which expression was conveyed from a choice of 7 labels (anger, disgust, fear, friendliness, happiness, meanness, sadness). Listeners' accuracy for each recording was noted as 0 or 1 to compute raw accuracy (Elfenbein et al. 2002; Scherer et al. 1991). To ensure that our recordings captured identifiable representations of the intended expressions, we performed one-sample *t* tests to compare listeners' accuracy to chance level, set as 1/number of alternatives in the forced-choice task (14.29%, or 1/7 labels). Further, to account for possible individual differences in listeners' detection thresholds, we also calculated listeners' accuracy using *d'*, an index of accuracy derived from signal detection theory (e.g., Macmillan and Creelman 1991) that represents the degree of overlap between theoretical distributions of signal and noise. A *d'* value of 1 would indicate that these distributions were 1 standard deviation apart, which would suggest fairly accurate identification of the stimuli (Spackman et al. 2009).

Participants' overall accuracy was 52.18% (children: 49.36%, adults: 54.18%), with 67.81% for anger (children: 62.01%, adults: 71.92%), 41.84% for disgust (children: 43.79%, adults: 40.47%), 50.16% for fear (children: 46.45%, adults: 52.79%), 56.13% for friendliness (children: 50.74%, adults: 59.94%), 37.32% for happiness (children: 37.32%, adults: 37.33%), 37.19% for meanness (children: 34.87%, adults: 38.84%), and 74.80% for sadness (children: 70.34%, adults: 77.97%). One-sample *t* tests revealed that listeners' accuracy rates for each expression were significantly higher than the chance level of 0.1429 (1/7 label options, or 14.29%; with anger, $t(146) = 44.27$, disgust, $t(146) = 23.98$, fear, $t(146) = 25.85$, friendliness, $t(146) = 30.76$, happiness, $t(146) = 19.89$, meanness, $t(146) = 16.77$, and sadness, $t(146) = 49.09$, all $ps < .001$). 94.29% of recordings (132/140) were recognized above chance level. Further, the *d'* value was 1.98 for anger (1.88 for child listeners, 2.05 for adult listeners), 1.08 for disgust (1.02 for children, 1.13 for adults), 1.54 for fear (1.34 for children, 1.69 for adults), 1.37 for friendliness (1.27 for children, 1.46 for adults), 1.44 for happiness (1.38 for children, 1.48 for adults), 1.12 for meanness (1.06 for children, 1.17 for adults), and 2.06 for sadness (1.88 for children, 2.20 for adults).

Our accuracy rates were slightly lower than those observed in previous studies, which have yielded an average accuracy of between 55% and 65% for adult listeners, depending on the type and number of emotion categories presented (Banse and Scherer 1996; Johnstone and Scherer 2000; van Bezooijen et al. 1983). For child listeners, reported values ranged from 52% to 72% in one investigation, depending on the participant's age (Sauter et al. 2012). However, these studies have tended to use fewer alternatives in their emotion recognition tasks, with an average of five label options (Juslin and Laukka 2003; Rosenthal and Rubin 1989). Compared to the seven labels used in the current investigation, such designs correspond to higher chance levels (20.0% versus 14.2%), which may be partially accounting for the lower accuracy in the current study. Several other design features could have impacted recognition rates. For one, our inclusion of a second positive expression (i.e., friendliness) along with happiness may have increased the task difficulty (Sauter and Scott 2007; Sauter et al. 2012). As well, our use of social expressions (i.e., meanness and friendliness), which are thought to be less intense than basic emotions (Mitchell and Ross 2013) and as such may be more ambiguous, could have reduced listeners' accuracy overall. Nonetheless, the patterns of recognition rates across expressions were consistent with previous findings (Banse and Scherer 1996; Iriondo et al. 2000; Johnstone and Scherer 2000; Juslin and Laukka 2001, 2003): notably, anger and sadness were among the best-recognized expressions, while disgust and happiness were poorly recognized by listeners. Thus, we are confident that our recordings captured prototypical patterns of vocal communication that were well-recognized by listeners.

Speech Analysis

Using Praat speech analysis software (Boersma and Weenink 2015), we extracted the following acoustic parameters from each recording.

Pitch Variables

The mean pitch in Hertz (Hz), its minimum point, and its maximum point were obtained over the whole sentence's waveform using autocorrelation, a time step of 0.0 s, pitch floor of 75 Hz, and a pitch ceiling of 600 Hz (to apply to higher pitched voices, like those of children). Pitch range was computed by subtracting the minimum pitch point from the maximum pitch point.

Intensity Variables

The mean energy intensity in decibels (dB), its minimum point, and its maximum point were obtained over the whole sentence's waveform, with a minimum pitch setting of 100 Hz and a time step of 0.0 s. Intensity range was computed by subtracting the minimum intensity point from the maximum intensity point.

Temporal Variable

Speech rate was computed by dividing the number of syllables in each sentence by the overall duration of the sentence, in seconds (Pell et al. 2009). This variable is thus expressed in syllables per second, where a higher value reflects a faster speech rate.

Measure of Voice Quality

Vocal quality measures serve a paralinguistic function (Gobl and Ní Chasaide 2000) and can improve listeners' recognition of the emotional intent of a speaker's voice when used in conjunction with other prosodic features (Zhang 2008). We extracted the mean harmonics-to-noise ratio (HNR) over the whole sentence. The ratio indexes the proportion of noise that interferes with the harmonics of the voice, and can be conceptualized as an index of 'hoarseness' (Yumoto et al. 1982). Indeed, HNR is related to listeners' ratings of breathiness and roughness, for both adults and youth (de Krom 1995; McAllister et al. 1996). A low ratio value represents more noise in the speech signal, leading to the impression of hoarseness. Conversely, a greater ratio value represents less noise in the speech signal, indicating a voice undisturbed by noise (Eskenazi et al. 1990). The mean harmonics-to-noise ratio was obtained using the Harmonicity function in Praat, with 0.01 s time steps, a minimum pitch of 75 Hz, silence threshold of 0.1, and 1 period per window.

The missing values from the problematic 90 recordings represented 2% of the full sample. We imputed this missing data by replacing the missing value for specific actors' recordings with the mean value for that specific recording (i.e., same emotion, sentence, and take) produced by other actors of the same age and gender (i.e., boys/girls/women/men).

Lastly, pitch and intensity variables were normalized for each speaker in reference to their resting frequency, following the procedure and formula outlined in Pell et al. (2009). For pitch, a resting frequency value was computed for each speaker, based on the average minimum pitch observed in all neutral sentences produced by that speaker. Normalized values of pitch mean were obtained with the following formula: $\text{PitchMean}_{\text{Normalized}} = (\text{PitchMean}_{\text{Observed}} - \text{Resting frequency})/\text{Resting frequency}$. For pitch range, we first normalized minimum pitch and maximum pitch in the same way, and subtracted the minimum value from the maximum value to obtain the range. The same procedure was used for intensity variables, using instead each speaker's average minimum intensity observed in all neutral sentences as that speaker's resting intensity value. The resulting normalized pitch and intensity variables thus represent their proportional distance from speakers' natural resting frequency (Pell et al. 2009). This procedure was done to control for differences between groups that may occur due to speaker-level differences in vocal tracts and natural frequency levels.

Statistical Analyses

We began by determining whether the extracted vocal cues could be meaningfully reduced to a smaller subset of factors. Given the hierarchical nature of the data (i.e., repeated observations nested within person), we examined both the within- and between-subject correlation matrices. As documented in Table 1, the magnitude of the within-subject correlations was generally small. The strongest association was between pitch mean and range, $r = .33$, $p < .001$; no other correlation exceeded .20. The between-subject correlations were stronger; there was a near-perfect correlation between intensity mean and intensity range, $r = .90$, $p < .001$, and a number of other associations had an absolute value greater than .45. In general, then, the pattern of results suggested meaningful associations of the investigated vocal cues between actors, but not across individual actors' performances. For example, actors who speak with higher mean intensity also show greater range of intensity. However, a change in actors' mean intensity across expressions was not associated with concomitant change in range of intensity, $r = -.05$, $p > .05$. As a second check on whether we should combine the vocal cues, we used MPlus 7.0 (Muthén and Muthén 2007) to perform a multilevel exploratory factor analysis (Reise et al. 2005), which accounts for the nesting of the data by returning both within- and between-factor solutions. This analysis did not return an interpretable solution. Based on these results, we decided to examine the vocal cues in separate analyses.

As such, repeated-measures analyses of variance (ANOVA) were performed to investigate the effect of the between-subject factors of age (2 levels) and gender (2 levels), and the within-subject factor of expression (7 levels: anger, disgust, fear, friendliness, happiness, meanness, sadness) on each of the acoustic cues listed above. Sentence (5 levels) was included as a covariate, given that vocal characteristics can interact with the grammatical features of a sentence (Scherer et al. 1984). Take (2 levels) was also included as a covariate. We performed the full-factorial model, including all main effects and

Table 1 Within- and between-subjects correlations of exploratory factor analysis model variables

Vocal cues					
Within-subjects	1	2	3	4	5
1. Pitch mean					
2. Pitch range	0.333***				
3. Intensity mean	0.141***	−0.095***			
4. Intensity range	0.088***	0.026	−0.048		
5. Mean HNR	0.137***	−0.105***	0.178***	0.153***	
6. Speech rate	−0.123***	−0.143***	0.006	−0.186***	0.037
Between-subjects					
1. Pitch mean					
2. Pitch range	0.519***				
3. Intensity mean	0.320**	−0.090			
4. Intensity range	0.264*	−0.146	0.904***		
5. Mean HNR	0.140	−0.537***	0.462***	0.474***	
6. Speech rate	−0.265*	0.001	−0.313**	−0.500***	−0.412***

This matrix contains results from the standardized model. * $p < .05$; ** $p < .01$; *** $p < .001$

interactions, for all cues separately. Greenhouse-Geisser corrections were applied when indicated by Mauchly's test of sphericity. Post-hoc Fisher's LSD tests and simple-effects tests were used to further examine significant effects.

Results

The mean values and 95% confidence intervals of our normalized measures of pitch and intensity variables, as well as of HNR and speech rate, as a function of expression, actor gender, and actor age, are displayed in Table 2. Results of the same analyses for additional vocal cues (duration, pitch min, pitch max, pitch variation, intensity min, intensity max, and intensity variation) can be obtained from the authors.

Given our research questions, we focus here on the effects of age and of expression; however, the results from the full-factorial model can be found in the "Appendix".

Main Effect of Expression

There was a main effect of expression for all variables under investigation (see "Appendix"). To establish whether our recordings' acoustic properties were consistent with previous findings on vocal paralanguage, we conducted Fisher's LSD post hoc comparisons for each variable. Given our interest in comparing meanness and friendliness to anger and happiness respectively, we will highlight these comparisons for each variable in turn.

For normalized pitch mean, post hoc comparisons demonstrated that all expressions were expressed with a higher pitch than neutral. Happiness, fear, and anger were the expressions that deviated most in pitch from neutral (*p*s < .05 compared to other expressions; *p*s > .05 compared to one another), and meanness deviated less from neutral than all other expressions (*p*s < .05). Specifically, meanness was significantly lower in

Table 2 Acoustic cues by expression, age group, and gender

Vocal cue	Expression	Average	Children			Adults		
		M [95% CI]	Girls ($n = 17$) *M* [95% CI]	Boys ($n = 7$) *M* [95% CI]	Average ($n = 24$) *M* [95% CI]	Women ($n = 15$) *M* [95% CI]	Men ($n = 15$) *M* [95% CI]	Average ($n = 30$) *M* [95% CI]
Pitch mean*	Anger	1.26 [1.10, 1.42]	1.33 [1.08, 1.58]	0.84 [0.45, 1.23]	1.08 [0.85, 1.32]	1.61 [1.34, 1.88]	1.25 [0.98, 1.52]	1.43 [1.23, 1.63]
	Disgust	0.98 [0.86, 1.10]	1.08 [0.88, 1.28]	0.64 [0.33, 0.95]	0.86 [0.67, 1.06]	1.15 [0.93, 1.37]	1.06 [0.84, 1.28]	1.10 [0.94, 1.26]
	Fear	1.34 [1.21, 1.48]	1.42 [1.17, 1.67]	0.94 [0.57, 1.31]	1.18 [0.94, 1.41]	1.79 [1.54, 2.04]	1.20 [0.95, 1.45]	1.49 [1.31, 1.67]
	Friendliness	1.09 [0.95, 1.23]	1.25 [1.03, 1.47]	0.81 [0.48, 1.14]	1.04 [0.84, 1.23]	1.32 [1.08, 1.56]	0.98 [0.74, 1.22]	1.15 [0.99, 1.31]
	Happiness	1.37 [1.23, 1.51]	1.43 [1.19, 1.67]	0.96 [0.59, 1.33]	1.20 [0.98, 1.41]	1.78 [1.53, 2.03]	1.30 [1.05, 1.55]	1.54 [1.36, 1.72]
	Meanness	0.88 [0.76, 1.00]	1.01 [0.81, 1.21]	0.59 [0.28, 0.90]	0.80 [0.63, 0.98]	1.05 [0.83, 1.27]	0.87 [0.65, 1.09]	0.96 [0.80, 1.12]
	Sadness	1.04 [0.90, 1.18]	1.19 [0.95, 1.43]	0.80 [0.45, 1.15]	1.00 [0.78, 1.21]	1.28 [1.03, 1.53]	0.88 [0.63, 1.13]	1.08 [0.90, 1.26]
Pitch range*	Anger	3.04 [2.82, 3.26]	2.73 [2.36, 3.10]	2.36 [1.77, 2.95]	2.55 [2.19, 2.90]	3.14 [2.75, 3.53]	3.91 [3.52, 4.30]	3.53 [3.26, 3.80]
	Disgust	3.24 [2.99, 3.49]	2.81 [2.38, 3.24]	2.45 [1.78, 3.12]	2.63 [2.24, 3.02]	3.23 [2.78, 3.68]	4.46 [4.01, 4.91]	3.85 [3.54, 4.16]
	Fear	3.02 [2.78, 3.26]	2.55 [2.16, 2.94]	2.27 [1.68, 2.86]	2.41 [2.05, 2.76]	3.18 [2.77, 3.59]	4.09 [3.68, 4.50]	3.63 [3.34, 3.92]
	Friendliness	3.12 [2.87, 3.37]	2.77 [2.36, 3.18]	2.33 [1.68, 2.98]	2.55 [2.16, 2.94]	3.23 [2.78, 3.68]	4.16 [3.71, 4.61]	3.69 [3.38, 4.00]
	Happiness	3.30 [3.06, 3.54]	2.76 [2.37, 3.15]	2.42 [1.81, 3.03]	2.59 [2.24, 2.94]	3.44 [3.03, 3.85]	4.58 [4.17, 4.99]	4.01 [3.72, 4.30]
	Meanness	3.06 [2.82, 3.30]	2.64 [2.27, 3.01]	2.44 [1.85, 3.03]	2.54 [2.19, 2.89]	3.07 [2.66, 3.48]	4.07 [3.66, 4.48]	3.57 [3.28, 3.86]
	Sadness	3.01 [2.79, 3.23]	2.47 [2.10, 2.84]	2.19 [1.60, 2.78]	2.33 [1.98, 2.68]	3.12 [2.73, 3.51]	4.26 [3.87, 4.65]	3.70 [3.43, 3.97]

Table 2 continued

Vocal cue	Expression	Average	Children			Adults		
		M [95% CI]	Girls (*n* = 17) *M* [95% CI]	Boys (*n* = 7) *M* [95% CI]	Average (*n* = 24) *M* [95% CI]	Women (*n* = 15) *M* [95% CI]	Men (*n* = 15) *M* [95% CI]	Average (*n* = 30) *M* [95% CI]
Int. mean*	Anger	0.62 [0.56, 0.68]	0.65 [0.57, 0.73]	0.52 [0.38, 0.66]	0.58 [0.50, 0.66]	0.75 [0.65, 0.85]	0.54 [0.44, 0.64]	0.65 [0.59, 0.71]
	Disgust	0.60 [0.54, 0.66]	0.64 [0.56, 0.72]	0.51 [0.37, 0.65]	0.58 [0.50, 0.66]	0.74 [0.64, 0.84]	0.50 [0.40, 0.60]	0.62 [0.56, 0.68]
	Fear	0.62 [0.56, 0.68]	0.66 [0.58, 0.74]	0.52 [0.38, 0.66]	0.59 [0.51, 0.67]	0.77 [0.67, 0.87]	0.53 [0.43, 0.63]	0.65 [0.59, 0.71]
	Friendliness	0.62 [0.58, 0.66]	0.65 [0.57, 0.73]	0.53 [0.41, 0.65]	0.59 [0.51, 0.67]	0.76 [0.68, 0.84]	0.53 [0.45, 0.61]	0.64 [0.58, 0.70]
	Happiness	0.62 [0.56, 0.68]	0.66 [0.58, 0.74]	0.52 [0.38, 0.66]	0.59 [0.51, 0.67]	0.76 [0.66, 0.86]	0.53 [0.43, 0.63]	0.64 [0.58, 0.70]
	Meanness	0.60 [0.54, 0.66]	0.63 [0.55, 0.71]	0.50 [0.38, 0.62]	0.57 [0.49, 0.65]	0.74 [0.66, 0.82]	0.51 [0.43, 0.59]	0.63 [0.57, 0.69]
	Sadness	0.61 [0.55, 0.67]	0.65 [0.57, 0.73]	0.52 [0.38, 0.66]	0.58 [0.50, 0.66]	0.75 [0.65, 0.85]	0.52 [0.42, 0.62]	0.63 [0.57, 0.69]
Int. range*	Anger	0.95 [0.89, 1.01]	1.04 [0.94, 1.14]	0.87 [0.73, 1.01]	0.96 [0.88, 1.03]	1.03 [0.93, 1.13]	0.84 [0.74, 0.94]	0.94 [0.86, 1.02]
	Disgust	0.89 [0.85, 0.93]	0.96 [0.88, 1.04]	0.79 [0.67, 0.91]	0.87 [0.79, 0.95]	1.02 [0.94, 1.10]	0.79 [0.71, 0.87]	0.91 [0.85, 0.97]
	Fear	0.90 [0.84, 0.96]	0.98 [0.88, 1.08]	0.83 [0.69, 0.97]	0.91 [0.83, 0.99]	0.99 [0.89, 1.09]	0.78 [0.68, 0.88]	0.89 [0.83, 0.95]
	Friendliness	0.83 [0.77, 0.89]	0.91 [0.83, 0.99]	0.74 [0.60, 0.88]	0.83 [0.75, 0.90]	0.95 [0.85, 1.05]	0.73 [0.63, 0.83]	0.84 [0.78, 0.90]
	Happiness	0.90 [0.84, 0.96]	0.96 [0.88, 1.04]	0.87 [0.73, 1.01]	0.92 [0.84, 0.99]	0.99 [0.89, 1.09]	0.77 [0.67, 0.87]	0.88 [0.82, 0.94]
	Meanness	0.89 [0.83, 0.95]	0.96 [0.86, 1.06]	0.82 [0.68, 0.96]	0.89 [0.81, 0.97]	1.00 [0.90, 1.10]	0.79 [0.69, 0.89]	0.90 [0.82, 0.98]
	Sadness	0.88 [0.82, 0.94]	0.93 [0.85, 1.01]	0.83 [0.69, 0.97]	0.88 [0.80, 0.96]	1.00 [0.90, 1.10]	0.78 [0.68, 0.88]	0.89 [0.83, 0.95]

Table 2 continued

Vocal cue	Expression	Average	Children			Adults		
		M [95% CI]	Girls ($n = 17$) *M* [95% CI]	Boys ($n = 7$) *M* [95% CI]	Average ($n = 24$) *M* [95% CI]	Women ($n = 15$) *M* [95% CI]	Men ($n = 15$) *M* [95% CI]	Average ($n = 30$) *M* [95% CI]
Speech rate	Anger	4.40 [4.18, 4.62]	4.08 [3.73, 4.43]	4.47 [3.92, 5.02]	4.28 [3.94, 4.61]	4.55 [4.18, 4.92]	4.49 [4.12, 4.86]	4.52 [4.27, 4.77]
	Disgust	4.39 [4.17, 4.61]	4.06 [3.71, 4.41]	4.71 [4.14, 5.28]	4.39 [4.05, 4.72]	4.25 [3.86, 4.64]	4.52 [4.13, 4.91]	4.39 [4.12, 4.66]
	Fear	4.71 [4.47, 4.95]	4.28 [3.89, 4.67]	4.86 [4.25, 5.47]	4.57 [4.19, 4.94]	4.73 [4.32, 5.14]	4.97 [4.56, 5.38]	4.85 [4.56, 5.14]
	Friendliness	5.09 [4.87, 5.31]	4.84 [4.49, 5.19]	5.00 [4.45, 5.55]	4.92 [4.58, 5.25]	5.06 [4.69, 5.43]	5.47 [5.10, 5.84]	5.26 [4.99, 5.53]
	Happiness	4.76 [4.54, 4.98]	4.44 [4.09, 4.79]	4.95 [4.38, 5.52]	4.69 [4.36, 5.03]	4.68 [4.29, 5.07]	5.00 [4.61, 5.39]	4.84 [4.57, 5.11]
	Meanness	4.72 [4.48, 4.96]	4.32 [3.91, 4.73]	4.80 [4.17, 5.43]	4.56 [4.19, 4.93]	4.72 [4.29, 5.15]	5.03 [4.60, 5.46]	4.88 [4.57, 5.19]
	Sadness	4.48 [4.26, 4.70]	4.14 [3.77, 4.51]	4.39 [3.82, 4.96]	4.26 [3.93, 4.60]	4.61 [4.22, 5.00]	4.78 [4.39, 5.17]	4.70 [4.43, 4.97]

Table 2 continued

Vocal cue	Expression	Average	Children			Adults		
		M [95% CI]	Girls ($n = 17$) *M* [95% CI]	Boys ($n = 7$) *M* [95% CI]	Average ($n = 24$) *M* [95% CI]	Women ($n = 15$) *M* [95% CI]	Men ($n = 15$) *M* [95% CI]	Average ($n = 30$) *M* [95% CI]
Mean HNR	Anger	12.80 [12.17, 13.43]	13.47 [12.43, 14.51]	13.98 [12.37, 15.59]	13.73 [12.77, 14.69]	13.48 [12.38, 14.58]	10.27 [9.17, 11.37]	11.87 [11.09, 12.65]
	Disgust	12.56 [11.91, 13.21]	13.29 [12.19, 14.39]	13.32 [11.61, 15.03]	13.31 [12.29, 14.33]	13.88 [12.70, 15.06]	9.74 [8.56, 10.92]	11.81 [10.99, 12.63]
	Fear	13.55 [12.92, 14.18]	14.39 [13.35, 15.43]	15.00 [13.37, 16.63]	14.70 [13.73, 15.66]	14.72 [13.60, 15.84]	10.11 [8.99, 11.23]	12.41 [11.63, 13.19]
	Friendliness	13.31 [12.68, 13.94]	13.51 [12.45, 14.57]	14.21 [12.56, 15.86]	13.86 [12.88, 14.84]	14.76 [13.64, 15.88]	10.77 [9.65, 11.89]	12.77 [11.97, 13.57]
	Happiness	13.53 [12.88, 14.18]	13.81 [12.73, 14.89]	14.68 [12.99, 16.37]	14.24 [13.24, 15.24]	14.72 [13.56, 15.88]	10.89 [9.73, 12.05]	12.81 [12.01, 13.61]
	Meanness	12.36 [11.71, 13.01]	12.77 [11.67, 13.87]	12.94 [11.23, 14.65]	12.86 [11.86, 13.85]	13.70 [12.54, 14.86]	10.03 [8.87, 11.19]	11.86 [11.04, 12.68]
	Sadness	13.63 [12.89, 14.37]	14.19 [12.97, 15.41]	15.57 [13.67, 17.47]	14.88 [13.74, 16.02]	14.55 [13.26, 15.84]	10.20 [8.91, 11.49]	12.37 [11.45, 13.29]

Mean values (*M*) of acoustic variables for each expression, by age group and gender. [95% CI] represents the 95% confidence interval. Int. = intensity, HNR = harmonicity-to-noise ratio. Speech rate is represented in syllables per second. Measures with an asterisk (*) have been normalized against a speaker's resting frequency or resting intensity in the neutral condition; the values for these vocal cues thus represent proportional distance from neutral normalized pitch mean than anger ($p < .001$), as was friendliness compared to happiness ($p < .001$).

Post hoc tests for normalized pitch range showed that happiness and disgust were expressed with greater deviation in pitch range from neutral (i.e., greater pitch range) than other expressions (and did not differ from one another, $p > .05$). All other expressions did not differ from one another (ps $> .05$). Specifically, meanness did not differ in pitch range from anger ($p > .05$), but friendliness was expressed with a smaller range of pitch than happiness ($p < .05$).

For normalized intensity mean, few differences between expressions were revealed with post hoc tests, though meanness was found to deviate least from neutral compared to other expressions (ps $< .05$, but did not differ from disgust, $p > .05$). Further, meanness was significantly lower in intensity mean than anger ($p < .05$), but friendliness did not differ from happiness ($p > .05$).

Post hoc tests on normalized intensity range demonstrated that anger had a greater range in intensity than all other expressions (all ps $< .05$) and that friendliness had the least range in intensity compared to other expressions (all ps $< .05$). Specifically, meanness was expressed with a smaller range in intensity than anger ($p < .05$), as was friendliness compared to happiness ($p < .05$).

For speech rate, post hoc comparisons revealed that friendliness was spoken with a faster speech rate than all other expressions, all ps $< .05$. Sadness, anger, and disgust were slowest in speech rate (ps $> .05$ with one another, but ps $< .05$ compared to other expressions). Specifically, meanness was faster in speech rate than anger ($p < .05$), as was friendliness compared to happiness ($p < .05$).

Lastly, post hoc tests on mean HNR showed that sadness, fear, happiness, and friendliness had the highest mean harmonics-to-noise ratio of all expressions (all ps $< .05$). Meanness was significantly lower in voice quality than all other expressions, all ps $< .05$, except disgust, from which it did not differ ($p > .05$). Specifically, meanness was lower in mean HNR than anger ($p < .05$), but friendliness did not differ from happiness ($p > .05$).

Effects of Age

No main effect of age was found for normalized pitch mean, but there was a significant interaction between age and expression, $F(4.73, 236.25) = 3.41$, $p = .01$, $\eta^2 = .06$ (see Fig. 2). Simple-effects tests revealed that the relative ordering of expressions in terms of normalized pitch mean was similar for both actor groups: for instance, happiness and fear differed most in pitch from neutral (and did not differ from one another, ps $> .05$), and meanness differed least (all ps $> .05$ compared to other expressions). However, the extent to which expressions differed from one another varied depending on actors' age. For adults, happiness, fear, and anger had the highest normalized pitch mean (and did not differ from one another, ps $> .05$), indicating that they were the most distinct from neutral. Friendliness, disgust, and sadness all had significantly lower normalized pitch mean than each of happiness, fear, and anger (ps $> .05$), and they did not differ from each other (ps $> .05$). Meanness had a lower normalized pitch mean than all other expressions (ps $< .05$). For children, the groups of expressions were less distinct: though happiness, fear, and anger were again most distinct from neutral in normalized pitch (ps $> .05$ from one another), friendliness did not differ in deviation from neutral from fear and anger (ps $> .05$), and sadness was comparable to friendliness and anger (ps $> .05$). Disgust and meanness were both least distinct from neutral in mean pitch (ps $> .05$). In general, adults

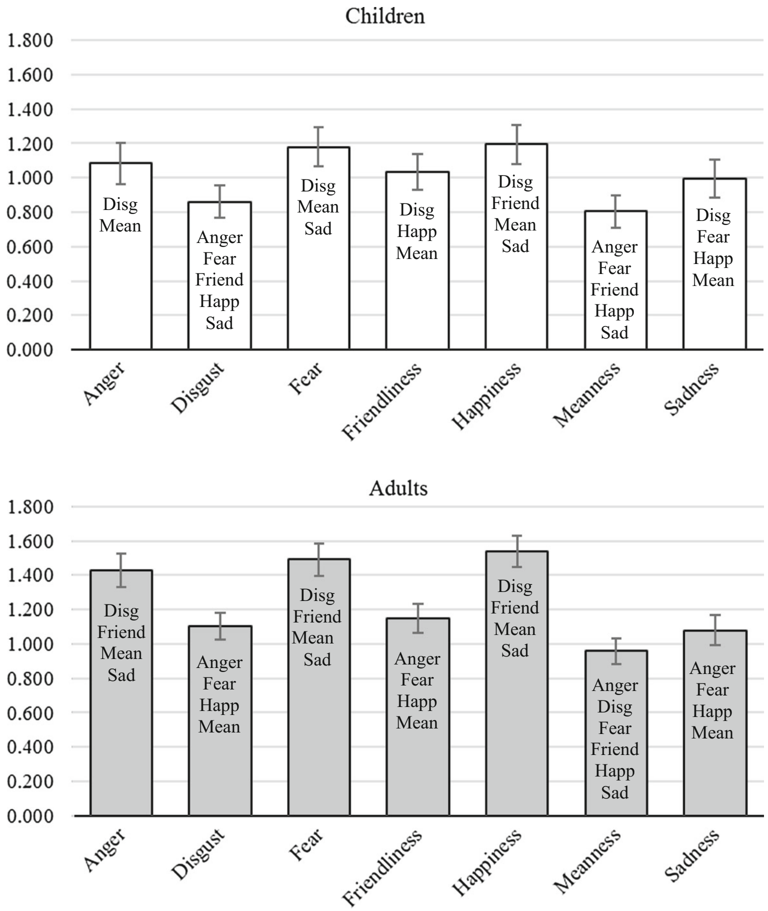

Fig. 2 Two-way interaction between expression and age on pitch mean (in proportional distance from neutral). Disg = disgust, Friend = friendliness, Happ = happiness, Mean = meanness, Sad = sadness. Simple-effects tests were performed to elucidate differences between expressions for each actor group. Pairwise comparisons were computed; letters on each bar represent the expressions from which that particular expression differs, *p*s < .05, for that actor group. A full report of all mean differences is available upon request

had more significant differences between expressions (30 significant differences out of 42 comparisons), compared to children (26 significant differences).

For normalized pitch range, there was a significant effect of age, $F(1, 50) = 33.98$, $p < .001$, $\eta^2 = .41$, qualified by an interaction between age and gender, $F(1, 50) = 10.72$, $p < .01$, $\eta^2 = .12$. Simple-effects tests suggested that adults of both genders deviated from neutral more than children, $F(1, 50) = 4.36$, $p < .05$ for women ($M = 3.20$) and girls ($M = 2.68$), and $F(1, 50) = 33.13$, $p < .001$ for men ($M = 4.22$) and boys ($M = 2.35$). Further, men had pitch ranges that deviated from neutral more than women, $F(1, 50) = 15.49$, $p < .001$. The age × expression interaction was not significant.

For mean HNR, there was a main effect of age, $F(1, 50) = 7.78$, $p < .01$, $\eta^2 = .14$, and a significant age × gender interaction, $F(1, 50) = 14.67$, $p < .001$, $\eta^2 = .23$. Simple-effects tests showed that girls ($M = 13.64$) and women ($M = 14.26$) did not differ in voice quality, but that boys ($M = 14.24$) had significantly higher voice quality than men ($M = 10.29$), $F(1, 50) = 17.51$, $p < .001$. Women also had higher voice quality than men, $F(1, 50) = 27.74$, $p < .001$. There was also an interaction between age and expression, $F(6, 300) = 3.39$, $p < .01$, $\eta^2 = .06$ (see Fig. 3). Simple-effects tests suggested that the relative ordering of expressions by mean HNR varied somewhat for children and adults. For instance, sadness and fear were highest in mean HNR for children (ps < .05), followed by happiness (which did not differ from fear, $p > .05$). For adults, the expressions with the highest mean HNR also included happiness and friendliness (ps > .05). Further, the distinctions between expressions varied for adults and children. For example, anger and meanness differed in HNR for children ($p < .05$), but not for adults ($p > .05$). Generally, a greater number of significant differences between expressions emerged for children (26 significant differences out of 42 possible comparisons) than for adults (18 of 42 comparisons).

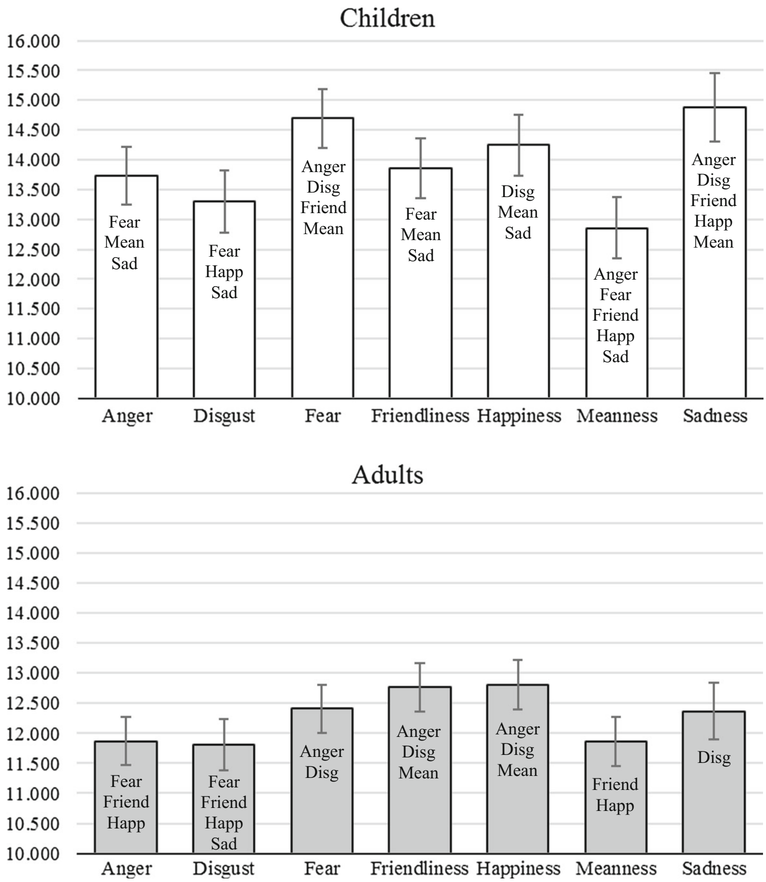

Fig. 3 Two-way interaction between expression and age on mean harmonics-to-noise ratio. Disg = disgust, Friend = friendliness, Happ = happiness, Mean = meanness, Sad = sadness. Simple-effects tests were performed to elucidate differences between expressions for each age group. Pairwise comparisons were computed; letters on each bar represent the expressions from which that particular expression differs, ps < .05, for that age group. A full report of all mean differences is available upon request

There were no main effects of age, interactions of age and expression, or interactions of age and gender for intensity mean, intensity range, or speech rate (see Appendix).

Discussion

We investigated differences in the acoustic cues underlying adults' and youth's portrayal of key socio-emotional expressions. Our primary goal was to examine whether emotional prosody varied as a function of age. Our results revealed age effects for pitch, which was normalized relative to speakers' resting frequency, and voice quality. First, adults deviated more in pitch range from baseline than children. This difference was larger between boys and men, but was also significant between girls and women. It is possible that adults' longer vocal tract (Fant et al. 1991; Harries et al. 1997; Hollien et al. 1994) may allow them to access a greater range of pitches when talking, compared to youth. If pitch is free to vary more for adults, this may enable them to represent various emotions using more distinct pitch profiles than children. Moreover, men's voices were lower in harmonics-to-noise ratios (i.e., more hoarse or rough-sounding; de Krom 1995) than children's and women's. In general, the literature on age and gender influences on voice quality is scarce and equivocal (Heffernan 2004; Stathopoulos et al. 2011). However, our results are consistent with some studies showing that elderly adults' voices contain more noise than younger adults' (Ferrand 2002; Gorham-Rowan and Laures-Gore 2006; c.f. Ferrand 2000) and that women have higher HNR values compared to men (Toran and Lal 2009; c.f. Klatt and Klatt 1990). Overall, our findings support the notion that developmental changes in laryngeal and respiratory systems can affect the production of vocal cues (Stathopoulos and Sapienza 1993). It will be important to investigate whether these age-related differences impact speakers' ability to convey socio-emotional information to listeners.

Recruiting both child and adult actors revealed that the effect of age varied as a function of expression for two indices: mean pitch and mean HNR. For pitch mean, the relative ordering of expressions, compared to one another and to neutral, was similar for both adults and children. However, there were significant differences in the degree to which each expression was distinguished from the others. Overall, adult speakers portrayed various expressions as more distinct in pitch from one another than children, for whom there was more frequent overlap in emotional categories. This may be due to the greater range of pitch values a longer adult vocal tract permits, as children have been found to demonstrate restricted vocal capabilities compared to typical adult values (McAllister et al. 1994). As well, relative to youth, adults' greater experience with communicating emotional intent to others over the lifespan may facilitate such differentiated portrayals. Pitch variables are crucial to the recognition of emotional intent (Pell et al. 2009; Scherer 1996); for this reason, the fact that adult actors portrayed expressions as more distinct in pitch than child actors may make their emotional paralanguage easier for listeners to interpret.

Moreover, we also found an interaction between age and expression for voice quality (mean HNR). Here, adults and children differed in the relative ordering of expressions, from high to low voice clarity, in comparison to neutral. For example, friendliness was conveyed with one of the highest values of HNR (i.e., clear-sounding voice) by adults, but not by children. Furthermore, children showed more distinctions between emotional categories than adults. This may not confer the same advantage in recognizability to children as pitch distinctions did to adults, since voice quality may be less integral to the interpretation of affect than pitch. However, this hypothesis must be investigated empirically.

Overall, we observed differences in the ways in which children and adults convey emotional prosody, suggesting that the two age groups may not be equivalent in their communication of socio-emotional information. For this reason, it is crucial to include both child- and adult-generated stimuli in cross-sectional investigations of the development of emotion recognition skills. Using expressions portrayed by adults only risks underestimating youth's ability to understand vocal cues. Youth may interact with their peers' emotional outputs more than those of adults, especially given display rules that constrain adults' expression of negative emotional states to children (e.g., Malatesta and Haviland 1982). Thus, youth's skills in interpreting adults' emotional stimuli may not generalize to their ability to decode other children's vocal affect. Given the current findings suggesting that adults' and children's vocal portrayals of socio-emotional information are not equivalent, it is crucial to include both types of stimuli in investigations of youth's ability to decode vocal affect. Moreover, age-related differences in the production of vocal emotional stimuli may impact listeners' ability to decode the intended emotion in speakers' voices. For instance, adult-produced expressions are more distinct in pitch, which may facilitate listeners' task of recognizing the intended expression in these recordings. Therefore, including stimuli produced by a variety of encoders may enable a systematic investigation of factors that impact listeners' emotion recognition skills, and thus improve our understanding of the communication of socio-emotional information.

Vocal Cues Associated with Meanness and Friendliness

The second objective of our study was to examine the acoustic correlates of meanness and friendliness, two important social expressions in children's social interactions, compared to other basic emotions, particularly anger and happiness. Analyses revealed significant main effects of expression for each of the vocal cues. Broadly, the acoustic characteristics associated with each basic emotion were consistent with known vocal profiles delineated in previous investigations; for example, happiness was expressed with high pitch mean and range, and sadness was portrayed with low pitch and slowed speech rate, as is typically reported (Banse and Scherer 1996; Bänziger et al. 2015; Iriondo et al. 2000; Johnstone and Scherer 2000; Juslin and Laukka 2001, 2003; Scherer 1996; Scherer and Oshinsky 1977; Ververidis and Kotropoulos 2006). Moreover, compared to other expressions, friendliness was expressed with a fast speech rate, low variation in intensity (relative to neutral), and high voice quality. Conversely, meanness was conveyed with low pitch and intensity mean (relative to neutral), and lower voice quality than other expressions. Beyond typifying the pattern of vocal cues associated with each, we also sought to determine whether these two expressions differed from other "basic" emotions, notably anger and happiness. As expected, friendliness was found to be significantly lower in pitch mean and range, but also faster in speech rate and lower in intensity range than happiness across all actor groups. Moreover, meanness was found to be expressed with lower pitch mean, lower intensity mean and range, faster speech rate, and lower voice quality than anger across all actor groups. These results suggest that the social expressions of meanness and friendliness are portrayed differently than their basic emotion counterparts of anger and happiness. Indeed, these expressions may be subtler than basic emotions, given that they were demarcated by less marked changes in pitch and intensity from baseline than were anger and happiness; this might make them more difficult for listeners to pick up on and interpret correctly. Overall, these findings are consistent with a growing body of evidence indicating that speakers' voices can convey more than basic emotions (Mitchell and Ross 2013), and can be an important source of information for social attitudes and intents.

Strengths and Limitations

Our study is the first to investigate the vocal cues underlying the communication of affective paralanguage in youth. Our results thus add to the literature on the development of vocal prosody by typifying these acoustic correlates and comparing them to those of adults. Further, we expand our knowledge about the communication of socio-affective information beyond basic emotions by investigating the vocal attributes of the social expressions ‘meanness’ and ‘friendliness’. We produced a sizeable collection of vocal affect portrayals, which incorporates a wide array of expressions enacted by a large sample of encoders; our findings thus describe the acoustic cues underlying a range of relevant social communications.

Our use of actors to portray the expressions under investigation in the current study, as opposed to obtaining recordings from non-actors or capturing spontaneous expressions in naturalistic settings, is both a strength and a limitation. Recruiting actors in studies on vocal communication is common (Scherer 2003), because it allows researchers to control for the influence of verbal content on prosody (Blaauw 1994; Howell and Kadi-Hanifi 1991). Moreover, the output of actors and non-actors has been found to be comparable perceptually (e.g., Jürgens et al. 2015). Nonetheless, recordings generated by actors in a laboratory may not be representative of naturalistic social exchanges. In particular, the child actors in this study may have provided prototypical exemplars of emotions that they learned from adults in acting classes or theatre programs. If their vocal representations are modeled on adults’, our design may be underestimating the differences between youth and adult actors’ affective prosody. Further, the recordings we obtained may not be representative of youth’s communication style in social interactions with peers, although outputs in peer exchanges may also be influenced by display rules and masking attempts that occur in naturalistic settings (Scherer et al. 1991). Nonetheless, despite the difficulties that arise in obtaining and analyzing spontaneous expressions, such designs are needed in future investigations to complement findings in laboratory settings.

Conclusion

The current study aimed to expand our understanding of the social information conveyed by the voice. First, we examined youth’s communication of emotional prosody in comparison to adults’. We determined that, despite some convergence in the vocal cues used by both age groups, there are marked differences in the acoustic characteristics underlying adults’ and children’s portrayals of various expressions. Such differences may be due to developmental changes in the physiological voice apparatus, but may also be related to adults’ greater experience in communicating varied and distinct emotional expressions over their lifespan, and to socialization influences that may modulate the ways in which we learn to convey affective information.

Overall, our findings have implications for the literature on emotion recognition. Given differences in adults’ and youth’s production of these socio-emotional cues, listeners’ ability to identify emotional intent may differ based on the speaker’s age. Indeed, using only adult-generated stimuli in investigations of children’s understanding of emotion may be underestimating youth’s ability to recognize socio-emotional cues. To understand fully the development of emotion recognition skills, researchers must include stimuli generated by encoders of different ages. Future studies should investigate whether characteristics of the speaker have a differential impact on listeners' emotion recognition abilities.

Further, our work extends understanding of the communication of social information, such as expressions of rejection (meanness) and affiliation (friendliness). Our results highlight the differences in the pattern of vocal cues used to convey these expressions, compared to similar basic emotions like anger and happiness. These findings suggest that expressions conveying attitudes and intents in social interactions should be included in investigations of emotion recognition to gain a comprehensive understanding of the communication of social information in interpersonal interactions.

Compliance with Ethical Standards

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

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the McGill Research Ethics Board, and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Appendix

See Table 3.

Table 3 ANOVA on the effect of age, gender, expression, sentence, and take on vocal cues

Vocal cues	*df*	*F*	η^2	*p*
Normalized pitch mean				
Age	(1, 50)	3.23	.06	.08
Gender	(1, 50)	9.89	.17	<.01**
Age × Gender	(1, 50)	0.16	<.01	.70
Expression	(4.73, 236.25)	38.47	.44	<.001***
Expression × Age	(4.73, 236.25)	3.41	.06	.01*
Expression × Gender	(4.73, 236.25)	2.40	.05	.04*
Expression × Age × Gender	(4.73, 222.08)	1.83	.04	.11
Take	(1, 50)	4.81	.09	.03*
Take × Age	(1, 50)	1.59	.03	.21
Take × Gender	(1, 50)	2.39	.05	.13
Take × Age × Gender	(1, 47)	2.31	.04	.14
Sentence	(3.28, 164.01)	31.31	.39	<.001***
Sentence × Age	(3.28, 164)	2.31	.04	.07
Sentence × Gender	(3.28, 164)	0.85	.02	.48
Sentence × Age × Gender	(3.28, 154.16)	0.30	.01	.84
Take × Sentence	(4, 200)	2.05	.04	.09
Take × Sentence × Age	(4, 164)	2.29	.04	.06
Take × Sentence × Gender	(4, 164)	1.38	.03	.24
Take × Sentence × Age × Gender	(4, 154.16)	0.94	.02	.44
Take × Expression	(4.71, 235.36)	2.63	.05	.03*
Take × Expression × Age	(4.71, 236.25)	0.86	.02	.50
Take × Expression × Gender	(4.71, 236.25)	1.51	.03	.19

Table 3 continued

Vocal cues	*df*	*F*	η^2	*p*
Take × Expression × Age × Gender	(4.71, 222.08)	1.29	.03	.27
Sentence × Expression	(12.93, 646.60)	5.29	.10	<.001***
Sentence × Expression × Age	(12.93, 774.90)	2.23	.04	.01*
Sentence × Expression × Gender	(12.93, 774.90)	0.72	.01	.74
Sentence × Expression × Age × Gender	(12.93, 728.41)	0.76	.02	.70
Take × Sentence × Expression	(13.76, 687.95)	0.49	.01	.94
Take × Sentence × Expression × Age	(13.76, 774.90)	1.06	.02	.40
Take × Sentence × Expression × Gender	(13.76, 774.90)	0.63	.01	.84
Take × Sentence × Expression × Age × Gender	(13.76, 728.41)	0.66	.01	.81
Normalized pitch range				
Age	(1, 50)	33.98	.41	<.001***
Gender	(1, 50)	2.86	.05	.10
Age × Gender	(1, 50)	10.72	.12	<.01**
Expression	(4.61, 230.33)	3.09	.06	<.05*
Expression × Age	(4.61, 230.25)	1.51	.03	.19
Expression × Gender	(4.61, 230.25)	0.56	.01	.72
Expression × Age × Gender	(4.61, 216.53)	0.42	.01	.82
Take	(1, 50)	0.12	<.01	.74
Take × Age	(1, 50)	0.23	.01	.63
Take × Gender	(1, 50)	3.90	.07	.05
Take × Age × Gender	(1, 47)	0.25	.01	.62
Sentence	(3.06, 152.81)	50.48	.50	<.001***
Sentence × Age	(3.06, 152.80)	1.10	.02	.35
Sentence × Gender	(3.06, 152.80)	4.63	.09	<.01**
Sentence × Age × Gender	(3.06, 143.63)	3.03	.06	.03*
Take × Sentence	(4, 200)	0.60	.01	.67
Take × Sentence × Age	(4, 152.80)	0.42	.01	.79
Take × Sentence × Gender	(4, 152.80)	0.38	.01	.82
Take × Sentence × Age × Gender	(4, 143.63)	0.33	.86	.01*
Take × Expression	(4.49, 224.61)	1.37	.03	.24
Take × Expression × Age	(4.49, 230.50)	0.18	<.01	.96
Take × Expression × Gender	(4.49, 230.50)	1.92	.04	.10
Take × Expression × Age × Gender	(4.49, 216.67)	0.11	<.01	.99
Sentence × Expression	(13.53, 676.65)	2.82	.05	<.001***
Sentence × Expression × Age	(13.53, 703.95)	0.95	.02	.50
Sentence × Expression × Gender	(13.53, 703.95)	1.11	.02	.35
Sentence × Expression × Age × Gender	(13.53, 662.14)	0.72	.01	.75
Take × Sentence × Expression	(12.72, 636.03)	0.72	.01	.74
Take × Sentence × Expression × Age	(12.72, 703.95)	0.92	.02	.53
Take × Sentence × Expression × Gender	(12.72, 703.95)	0.97	.02	.48
Take × Sentence × Expression × Age × Gender	(12.72, 661.71)	1.17	.02	.30
Normalized intensity mean				
Age	(1, 50)	1.19	.02	.28

Table 3 continued

Vocal cues	*df*	*F*	η^2	*p*
Gender	(1, 50)	13.38	.21	.001**
Age × Gender	(1, 50)	0.97	.02	.33
Expression	(6, 300)	6.45	.11	<.001***
Expression × Age	(6, 300)	0.82	.02	.55
Expression × Gender	(6, 300)	1.21	.02	.30
Expression × Age × Gender	(6, 282)	0.52	.01	.79
Take	(1, 50)	4.71	.09	.04*
Take × Age	(1, 50)	1.38	.03	.25
Take × Gender	(1, 50)	0.58	.01	.45
Take × Age × Gender	(1, 47)	0.48	.01	.49
Sentence	(3.27, 163.53)	32.51	.39	<.001***
Sentence × Age	(3.27, 163.55)	0.60	.01	.63
Sentence × Gender	(3.27, 163.55)	1.78	.03	.15
Sentence × Age × Gender	(3.27, 153.74)	0.70	.01	.57
Take × Sentence	(4, 200)	2.41	.05	.05
Take × Sentence × Age	(4, 163.55)	0.69	.01	.60
Take × Sentence × Gender	(4, 163.55)	0.38	.01	.82
Take × Sentence × Age × Gender	(4, 153.74)	1.70	.03	.15
Take × Expression	(6, 300)	2.03	.04	.06
Take × Expression × Age	(6, 300)	0.42	.01	.87
Take × Expression × Gender	(6, 300)	1.29	.03	.26
Take × Expression × Age × Gender	(6, 282)	1.97	.04	.07
Sentence × Expression	(14.48, 723.97)	2.01	.04	.01*
Sentence × Expression × Age	(14.48, 981.30)	1.30	.03	.20
Sentence × Expression × Gender	(14.48, 981.30)	1.23	.02	.25
Sentence × Expression × Age × Gender	(14.48, 922.42)	0.53	.01	.92
Take × Sentence × Expression	(13.84, 691.98)	0.51	.01	.93
Take × Sentence × Expression × Age	(13.84, 981.30)	1.58	.03	.08
Take × Sentence × Expression × Gender	(13.84, 981.30)	1.13	.02	.33
Take × Sentence × Expression × Age × Gender	(13.84, 922.42)	1.06	.02	.39
Normalized intensity range				
Age	(1, 50)	0.01	<.001	.94
Gender	(1, 50)	12.61	.20	<.01**
Age × Gender	(1, 50)	0.53	.01	.47
Expression	(6, 300)	15.60	.24	<.001***
Expression × Age	(6, 300)	2.01	.04	.06
Expression × Gender	(6, 300)	0.76	.02	.60
Expression × Age × Gender	(6, 282)	1.27	.03	.27
Take	(1, 50)	0.26	.01	.61
Take × Age	(1, 50)	0.68	.01	.41
Take × Gender	(1, 50)	0.14	<.01	.71
Take × Age × Gender	(1, 47)	0.95	.02	.33
Sentence	(3.27, 163.58)	43.48	.47	<.001***

Table 3 continued

Vocal cues	*df*	*F*	η^2	*p*
Sentence × Age	(3.27, 163.60)	0.90	.02	.45
Sentence × Gender	(3.27, 163.60)	1.30	.03	.28
Sentence × Age × Gender	(3.27, 153.78)	3.72	.07	.01*
Take × Sentence	(4, 200)	3.13	.06	.02*
Take × Sentence × Age	(4, 163.60)	1.14	.02	.34
Take × Sentence × Gender	(4, 163.60)	0.31	.01	.87
Take × Sentence × Age × Gender	(4, 153.78)	1.40	.03	.24
Take × Expression	(6, 300)	1.05	.02	.39
Take × Expression × Age	(6, 300)	1.01	.02	.42
Take × Expression × Gender	(6, 300)	1.84	.04	.09
Take × Expression × Age × Gender	(6, 282)	1.52	.03	.17
Sentence × Expression	(24, 1200)	2.02	.04	<.01**
Sentence × Expression × Age	(24, 981.60)	0.81	.02	.72
Sentence × Expression × Gender	(24, 981.60)	0.66	.01	.89
Sentence × Expression × Age × Gender	(24, 922.70)	0.85	.02	.68
Take × Sentence × Expression	(14.94, 747.09)	0.54	.01	.92
Take × Sentence × Expression × Age	(14.94, 981.60)	1.18	.02	.28
Take × Sentence × Expression × Gender	(14.94, 981.60)	1.27	.03	.21
Take × Sentence × Expression × Age × Gender	(14.94, 922.70)	0.49	.01	.95
Speech rate				
Age	(1, 50)	1.53	.03	.22
Gender	(1, 50)	2.70	.05	.11
Age x Gender	(1, 50)	0.23	.01	.63
Expression	(4.80, 239.84)	23.90	.32	<.001***
Expression × Age	(4.80, 239.85)	1.93	.04	.09
Expression × Gender	(4.80, 239.85)	1.20	.02	.31
Expression × Age × Gender	(4.80, 225.46)	1.30	.03	.26
Take	(1, 50)	0.03	<.01	.87
Take × Age	(1, 50)	3.93	.07	.05
Take × Gender	(1, 50)	2.71	.05	.11
Take × Age × Gender	(1, 47)	1.91	.04	.17
Sentence	(4, 200)	221.22	.82	<.001***
Sentence × Age	(4, 200)	7.10	.12	<.001***
Sentence × Gender	(4, 200)	1.58	.03	.18
Sentence × Age × Gender	(4, 188)	4.09	.08	<.01**
Take × Sentence	(4, 200)	0.80	.02	.53
Take × Sentence × Age	(4, 200)	2.27	.04	.06
Take × Sentence × Gender	(4, 200)	0.89	.02	.47
Take × Sentence × Age × Gender	(4, 188)	1.03	.02	.39
Take × Expression	(4.51, 225.54)	3.06	.06	.01*
Take × Expression × Age	(4.51, 239.85)	0.88	.02	.49
Take × Expression × Gender	(4.51, 239.85)	1.83	.04	.12
Take × Expression × Age × Gender	(4.51, 225.46)	0.75	.02	.57

Table 3 continued

Vocal cues	*df*	*F*	η^2	*p*
Sentence × Expression	(13.69, 684.24)	3.75	.07	<.001***
Sentence × Expression × Age	(13.69, 959.40)	2.31	.04	.01*
Sentence × Expression × Gender	(13.69, 959.40)	0.89	.02	.57
Sentence × Expression × Age × Gender	(13.69, 901.84)	1.29	.03	.21
Take × Sentence × Expression	(11.84, 592.02)	1.40	.03	.16
Take × Sentence × Expression × Age	(11.84, 959.40)	1.89	.04	.03*
Take × Sentence × Expression × Gender	(11.84, 959.40)	1.33	.03	.20
Take × Sentence × Expression × Age × Gender	(11.84, 901.84)	1.23	.02	.26
Mean harmonics-to-noise ratio				
Age	(1, 50)	7.78	.14	<.01**
Gender	(1, 50)	7.92	.14	<.01**
Age × Gender	(1, 50)	14.67	.23	<.001***
Expression	(6, 300)	11.25	.18	<.001***
Expression × Age	(6, 300)	3.39	.06	<.01**
Expression × Gender	(6, 300)	0.76	.02	.60
Expression × Age × Gender	(6, 282)	1.37	.03	.23
Take	(1, 50)	0.16	<.01	.69
Take × Age	(1, 50)	0.38	.01	.54
Take × Gender	(1, 50)	<0.001	<.001	>.99
Take × Age × Gender	(1, 47)	1.14	.02	.29
Sentence	(4, 200)	18.46	.27	<.001***
Sentence × Age	(4, 200)	2.38	.05	.05
Sentence × Gender	(4, 200)	1.87	.04	.12
Sentence × Age × Gender	(4, 188)	3.77	.07	.01*
Take × Sentence	(4, 200)	3.78	.07	.01*
Take × Sentence × Age	(4, 200)	3.22	.06	.01*
Take × Sentence × Gender	(4, 200)	3.78	.07	.01*
Take × Sentence × Age × Gender	(4, 188)	1.12	.02	.35
Take × Expression	(4.54, 226.82)	2.31	.04	.05
Take × Expression × Age	(4.54, 300)	2.15	.04	.07
Take × Expression × Gender	(4.54, 300)	0.31	.01	.90
Take × Expression × Age × Gender	(4.54, 282)	0.70	.01	.61
Sentence × Expression	(15.32, 765.74)	3.05	.06	<0.01***
Sentence × Expression × Age	(15.32, 1200)	1.41	.03	.14
Sentence × Expression × Gender	(15.32, 1200)	0.83	.02	.65
Sentence × Expression × Age × Gender	(15.32, 1128)	0.74	.02	.75
Take × Sentence × Expression	(13.51, 675.34)	0.69	.01	.79
Take × Sentence × Expression × Age	(13.51, 1200)	0.81	.02	.66
Take × Sentence × Expression × Gender	(13.51, 1200)	0.80	.02	.67
Take × Sentence × Expression × Age × Gender	(13.51, 1128)	0.86	.02	.60

References

Allgood, R., & Heaton, P. (2015). Developmental change and cross-domain links in vocal and musical emotion recognition performance in childhood. *British Journal of Developmental Psychology, 33,* 398–403.

Averill, J. R. (1983). Studies on anger and aggression: Implications for theories of emotion. *American Psychologist, 38,* 1145–1160.

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

Bänziger, T., Hosoya, G., & Scherer, K. R. (2015). Path models of vocal emotion communication. *PLoS ONE, 10,* 1–29.

Bayes, M. A. (1972). Behavioral cues of interpersonal warmth. *Journal of Counseling and Clinical Psychology, 39,* 333–339.

Blaauw, E. (1994). The contribution of prosodic boundary markers to the perceptual difference between read and spontaneous speech. *Speech Communication, 14,* 359–375.

Boersma, P., & Weenink, D. (2015). Praat: Doing phonetics by computer [Computer program]. Version 5.4.09, retrieved 1 June 2015 from http://www.praat.org/.

Brosgole, L., & Weisman, J. (1995). Mood recognition across the ages. *International Journal of Neuroscience, 82,* 169–189.

Cheang, H. S., & Pell, M. D. (2008). The sound of sarcasm. *Speech Communication, 50,* 366–381.

Chronaki, G., Hadwin, J. A., Garner, M., Maurage, P., & Sonuga-Barke, E. J. S. (2014). The development of emotion recognition from facial expressions and non-linguistic vocalizations during childhood. *British Journal of Developmental Psychology, 33,* 218–236.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115,* 74–101.

Cunningham, M. R. (1988). Does happiness mean friendliness? Induced mood and heterosexual self-disclosure. *Personality and Social Psychology Bulletin, 14,* 283–297.

de Krom, G. (1995). Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments. *Journal of Speech and Hearing Research, 38,* 794–811.

Dodge, K. A., & Coie, J. D. (1987). Social-information-processing factors in reactive and proactive aggression in children's peer groups. *Journal of Personality and Social Psychology, 53,* 1146–1158.

Ekman, P., Friesen, W. V., & Ancoli, S. (1980). Facial signs of emotional experience. *Journal of Personality and Social Psychology, 39,* 1125–1134.

Elfenbein, H. A., Mandal, M. K., Ambady, N., Harizuka, S., & Kumar, S. (2002). Cross-cultural patterns in emotion recognition: Highlighting design and analytical techniques. *Emotion, 2,* 75–84.

Eskenazi, L., Childers, D. G., & Hicks, D. M. (1990). Acoustic characteristics of vocal quality. *Journal of Speech, Language, and Hearing Research, 33,* 298–306.

Eskritt, M., & Lee, K. (2003). Do actions speak louder than words? Preschool children's use of the verbal-nonverbal consistency principle during inconsistent communications. *Journal of Nonverbal Behavior, 27,* 25–41.

Fant, G., Kruckenberg, A., & Nord, L. (1991). Prosodic and segmental speaker variations. *Speech Communication, 10,* 521–531.

Ferrand, C. T. (2000). Harmonics-to-noise ratios in normally speaking prepubescent girls and boys. *Journal of Voice, 14,* 17–21.

Ferrand, C. T. (2002). Harmonics-to-noise ratio: An index of vocal aging. *Journal of Voice, 16,* 480–487.

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *Journal of the Acoustical Society of America, 106,* 1511–1522.

Fives, C. J., Kong, G., Fuller, J. R., & DiGiuseppe, R. (2011). Anger, aggression, and irrational beliefs in adolescents. *Cognitive Therapy and Research, 35,* 199–208.

Friend, M. (2000). Developmental changes in sensitivity to vocal paralanguage. *Developmental Science, 3,* 148–162.

Gobl, C., & Ní Chasaide, A. (2000). Testing affective correlates of voice quality through analysis and resynthesis. In *ISCA tutorial and research workshop (ITRW) on Speech and Emotion.*

Gorham-Rowan, M. M., & Laures-Gore, J. (2006). Acoustic-perceptual correlates of voice quality in elderly men and women. *Journal of Communication Disorders, 39,* 171–184.

Gross, J. J., Carstensen, L. L., Pasupathi, M., Tsai, J., Götestam Skorpen, C., & Hsu, A. Y. C. (1997). Emotion and aging: Experience, expression, and control. *Psychology and Aging, 12,* 590–599.

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

Harries, M. L. L., Walker, J. M., Williams, D. M., Hawkins, S., & Hughes, I. A. (1997). Changes in the male voice at puberty. *Archives of Disease in Childhood, 77,* 445–447.
Heffernan, K. (2004). Evidence from HNR that /s/ is a social marker of gender. *Toronto Working Papers in Linguistics, 23*, 71–84.
Hollien, H., Green, R., & Massey, K. (1994). Longitudinal research on adolescent voice change in males. *Journal of the Acoustical Society of America, 96,* 2646–2654.
Hortensius, R., Schutter, D. J. L. G., & Harmon-Jones, E. (2012). When anger leads to aggression: Induction of relative left frontal cortical activity with transcranial direct current stimulation increases the anger-aggression relationship. *Social Cognitive and Affective Neuroscience, 7,* 342–347.
Howell, P., & Kadi-Hanifi, K. (1991). Comparison of prosodic properties between read and spontaneous speech material. *Speech Communication, 10,* 163–169.
Hubbard, J. A., Smithmyer, C. M., Ramsden, S. R., Parker, E. H., Flanagan, K. D., Dearing, K. F., et al. (2002). Observational, physiological, and self-report measures of children's anger: Relations to reactive versus proactive aggression. *Child Development, 73,* 1101–1118.
Iriondo, I., Guaus, R., Rodriguez, A., Lazaro, P., Montoya, N., Blanco, J. M., Bernadas, D., Oliver, J. M., Tena, D., & Longhi, L. (2000). Validation of an acoustical modeling of emotional expression in Spanish using speech synthesis techniques. In *Proceedings of the ISCA ITRW on speech and emotion, Newcastle, 5–7 September, 2000* (pp. 161–166). Belfast: Textflow.
Izard, C., Fine, S., Schultz, D., Mostow, A., Ackerman, B., & Youngstrom, E. (2001). Emotion knowledge as a predictor of social behaviour and academic competence in children at risk. *Psychological Science, 12,* 18–23.
Johnstone, T., & Scherer, K. R. (2000). Vocal communication of emotion. In M. Lewis & J. Haviland (Eds.), *The handbook of emotion* (pp. 220–235). New York: Guilford.
Jürgens, R., Grass, A., Drolet, M., & Fischer, J. (2015). Effect of acting experience on emotion expression and recognition in voice: Non-actors provide better stimuli than expected. *Journal of Nonverbal Behaviour, 39,* 195–214.
Juslin, P. N., & Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion, 1,* 381–412.
Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129,* 770–814.
Keltner, D., & Ekman, P. (2000). Facial expression of emotion. In M. Lewis & J. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 236–249). New York: Guilford Publications.
Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America, 87,* 820–857.
LaFrance, M. (2002). Smile boycotts and other body politics. *Feminism & Psychology, 12,* 319–323.
Laukka, P., Linnman, C., Ahs, F., Pissiota, A., Frans, O., Faria, V., et al. (2008). In a nervous voice: Acoustic analysis and perception of anxiety in social phobics' speech. *Journal of Nonverbal Behaviour, 32,* 195–214.
Leinonen, L., Hiltunen, T., Linnankoski, I., & Laakso, M. L. (1997). Expression of emotional-motivational connotations with a one-word utterance. *Journal of the Acoustical Society of America, 102,* 1853–1863.
Leppänen, J. M., & Hietanen, J. K. (2001). Emotion recognition and social adjustment in school-aged girls and boys. *Scandinavian Journal of Psychology, 42,* 429–435.
Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.
Malatesta, C. Z., & Haviland, J. M. (1982). Learning display rules: The socialization of emotion expression in infancy. *Child Development, 53,* 991–1003.
Masten, C. L., Eisenberger, N. I., Borofsky, L. A., Pfeifer, J. H., McNealy, K., Mazziotta, J. C., et al. (2009). Neural correlates of social exclusion during adolescence: Understanding the distress of peer rejection. *Social Cognitive and Affective Neuroscience, 4,* 143–157.
Maxim, L. A., & Nowicki, S. J. (2003). Developmental associations between nonverbal ability and social competence. *Philosophy, Sociology and Psychology, 2,* 745–758.
McAllister, A., Sederholm, E., Sundberg, J., & Gramming, P. (1994). Relations between voice range profiles and physiological and perceptual voice characteristics in ten-year-old children. *Journal of Voice, 8,* 230–239.
McAllister, A., Sundberg, J., & Hibi, S. R. (1996). Acoustic measurements and perceptual evaluation of hoarseness in children's voices. *Speech, Music and Hearing Quarterly Progress and Status Report, 4,* 15–26.
Mitchell, R. L. C., & Ross, E. D. (2013). Attitudinal prosody: What we know and directions for future study. *Neuroscience and Biobehavioral Reviews, 37,* 471–479.
Muthén, L. K., & Muthén, B. O. (2007). *Mplus [Software].* Los Angeles, CA: Muthén & Muthén.

Noble, L., & Xu, Y. (2011). Friendly speech and happy speech—are they the same? *Proceedings of the 17th international congress of phonetic sciences*. Hong Kong.

Paquette, J. A., & Underwood, M. K. (1999). Gender differences in young adolescents' experiences of peer victimization: Social and physical aggression. *Merrill-Palmer Quarterly, 45,* 242–266.

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

Petersen, A. C., Crockett, L., Richards, M., & Boxer, A. (1988). A self-report measure of pubertal status: Reliability, validity and initial norms. *Journal of Youth and Adolescence, 17,* 117–133.

Pittam, J., & Scherer, K. R. (1993). Vocal expression and communication of emotion. In M. Lewis & J. M. Haviland (Eds.), *Handbook of emotions* (pp. 185–197). New York, London: The Guilford Press.

Pons, F., & Harris, P. (2005). Longitudinal change and longitudinal stability of individual differences in children's emotion understanding. *Cognition and Emotion, 19,* 1158–1174.

Reise, S. P., Ventura, J., Nuechterlein, K. H., & Kim, K. H. (2005). An illustration of multilevel factor analysis. *Journal of Personality Assessment, 84,* 126–136.

Rigby, K. (2000). Effects of peer victimization in schools and perceived social support on adolescent well-being. *Journal of Adolescence, 23,* 57–68.

Rochman, D., Diamond, G. M., & Amir, O. (2008). Unresolved anger and sadness: Identifying vocal acoustical correlates. *Journal of Counseling Psychology, 55,* 505–517.

Rockwell, P. (2000). Lower, slower, louder: Vocal cues of sarcasm. *Journal of Psycholinguistic Research, 29,* 483–495.

Rosenthal, R., & Rubin, D. B. (1989). Effect size estimation for one-sample multiple-choice-type data: Design, analysis, and meta-analysis. *Psychological Bulletin, 106,* 332–337.

Sauter, D. A., Panattoni, C., & Happé, F. (2012). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology, 31,* 97–113.

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

Scherer, K. R. (1996). Adding the affective dimension: A new look in speech analysis and synthesis. *Proceedings of the 4th international conference of spoken language processing*. Philadelphia, USA.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40,* 227–256.

Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion, 15,* 123–148.

Scherer, K. R., Koivumaki, J., & Rosenthal, R. (1972). Minimal cues in the vocal communication of affect: Judging emotions from content-masked speech. *Journal of Psycholinguistic Research, 1,* 269–285.

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

Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion, 1,* 331–346.

Simmons, R. (2002). *Odd girl out: The hidden culture of aggression in girls*. New York: Harcourt.

Spackman, M. P., Brown, B. L., & Otto, S. (2009). Do emotions have distinct vocal profiles? A study of idiographic patterns of expression. *Cognition and Emotion, 23,* 1565–1588.

Stathopoulos, E. T., Huber, J. E., & Sussman, J. E. (2011). Changes in acoustic characteristics of the voice across the life span: Measures from individuals 4–93 years of age. *Journal of Speech, Language, and Hearing Research, 54,* 1011–1021.

Stathopoulos, E. T., & Sapienza, C. (1993). Respiratory and laryngeal measures of children during vocal intensity variation. *Journal of the Acoustical Society of America, 94,* 2531–2543.

Sussman, J. E., & Sapienza, C. (1994). Articulatory, developmental, and gender effects on measures of fundamental frequency and jitter. *Journal of Voice, 8,* 145–156.

Toran, K. C., & Lal, B. K. (2009). Objective analysis of voice in normal young adults. *Kathmandu University Medical Journal, 7,* 374–377.

Trentacosta, C. J., & Fine, S. E. (2010). Emotion knowledge, social competence, and behavior problems in childhood and adolescence: A meta-analytic review. *Social Development, 19,* 1–29.

Underwood, M. K. (2003). *Social aggression among girls*. New York: Guilford.

Underwood, M. K. (2004). Glares of contempt, eye rolls of disgust and turning away to exclude: Non-verbal forms of social aggression among girls. *Feminism Psychology, 14,* 371–375.

van Bezooijen, R., Otto, S. A., & Heenan, T. A. (1983). Recognition of vocal expressions of emotion: A three-nation study to identify universal characteristics. *Journal of Cross-Cultural Psychology, 14,* 387–406.

Ververidis, D., & Kotropoulos, C. (2006). Emotional speech recognition: Resources, features, and methods. *Speech Communication, 48,* 1162–1181.

Whiteside, S. P., & Hodgson, C. (2000). Speech patterns of children and adults elicited via a picture-naming task: An acoustic study. *Speech Communication, 32,* 267–285.
Wilkowski, B. M., & Robinson, M. D. (2008). The cognitive basis of trait anger and reactive aggression: An integrative analysis. *Personality and Social Psychology Review, 12,* 3–21.
Yumoto, E., Gould, W. J., & Baer, T. (1982). Harmonics-to-noise ratio as an index of the degree of hoarseness. *Journal of the Acoustical Society of America, 71,* 1544–1550.
Zhang, S. (2008). Emotion recognition in Chinese natural speech by combining prosody and voice quality features. *Advances in Neural Networks, 5264,* 457–464.
Zupan, B. (2015). Recognition of high and low intensity facial and vocal expressions of emotion by children and adults. *Journal of Social Sciences and Humanities, 1,* 332–344.