

Mid-Adolescents' and Adults' Recognition of Vocal Cues of Emotion and Social Intent: Differences by Expression and Speaker Age

Michele Morningstar¹ · Verity Y. Ly¹ · Lara Feldman¹ ·
Melanie A. Dirks¹

Published online: 16 January 2018

© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract Previous research has suggested that the ability to recognize vocal portrayals of socio-emotional expressions improves with age throughout childhood and adolescence. The current study examined whether stimulus-level factors (i.e., the age of the speaker and the type of expression being conveyed) interacted with listeners' developmental stage to predict listeners' recognition accuracy. We assessed mid-adolescent ($n = 50$, aged 13–15 years) and adult ($n = 87$, 18–30 years) listeners' ability to recognize basic emotions and social expressions in the voices of both adult and youth actors. Adults' emotional prosody was better recognized than that of youth, and adult listeners were more accurate overall than were mid-adolescents. Interaction effects revealed that youths' accuracy was equivalent to adult listeners' when hearing adult portrayals of anger, disgust, friendliness, happiness, and meanness, and youth portrayals of disgust, happiness, and meanness. Our findings highlight the importance of speaker characteristics and type of expression on listeners' ability to recognize vocal cues of emotion and social intent.

Keywords Emotion recognition · Social expressions · Vocal cues · Adolescence

Introduction

The human voice is an important source of social information, as vocal cues can convey a speaker's emotional state (Banse and Scherer 1996) and social intentions (Mitchell and Ross 2013) beyond the content of their speech (Scherer et al. 1991). The ability to decode socio-emotional cues from the voice has been found to increase with age until early to mid-adolescence, at which time listeners begin to show adult-like levels of skill (Brosgole and Weisman 1995; Zupan 2015). However, little is known about whether these age-related

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

¹ Department of Psychology, McGill University, 2001 McGill College, Room 1401, Montreal, QC H3A 1G1, Canada

changes vary as a function of stimulus properties, such as the age of the speaker. Being able to recognize the vocal cues of their peers is critical for youths' social interactions, and recent evidence indicates that the emotional prosody of youth may not be equivalent to that of adults (Morningstar et al. 2017a). The current study investigated mid-adolescent and adult listeners' recognition of vocal socio-emotional expressions portrayed by both youth and adults. We examined whether the developmental stage of the listener interacted with stimulus-level properties—specifically, the speaker's age and the type of expression being conveyed—to predict listeners' vocal emotion recognition (ER) ability.

The development of vocal ER skills is thought to be supported by age-related maturation of neural structures and circuits involved in the processing of affective information (Blakemore and Choudhury 2006; Blasi et al. 2011; Grossman 2010; Grossmann et al. 2010), coupled with increased cognitive abilities that allow for the appropriate appraisal of emotion cues (e.g., Nelson et al. 2005). Indeed, it has been documented that the ability to recognize vocal expressions of emotion improves throughout childhood (e.g., Allgood and Heaton 2015; Chronaki et al. 2015; Sauter et al. 2012), until reaching adult-like levels in early to mid-adolescence (Brosigole and Weisman 1995; Zupan 2015). However, recent work demonstrating continued maturation of the “social brain network” in adolescence (see Kilford et al. 2016 for a review) suggests that additional fine-tuning of emotional processes may be occurring until adulthood. Thus, adolescents' ER skills may be continuing to improve in subtle ways, and their performance relative to adults may depend on the complexity of the stimulus. Indeed, Brunswik's lens model of dyadic communication (Brunswik 1956) stipulates that properties of the stimulus could impact listeners' decoding accuracy, or interact with developmental influences to hinder or facilitate recognition of emotional prosody. The current study examines whether age-related changes in vocal ER skills may vary as a function of two stimulus properties: speaker age and expression type.

Speaker Age

The age of the speaker may be linked to listeners' ER accuracy. Though the Diagnostic Analysis of Nonverbal Accuracy (DANVA-2; Nowicki and Duke 2001; Rothman and Nowicki 2004) contains subtests assessing vocal ER in response to both child and adult-generated stimuli, work using this measure has not yet systematically compared the recognizability of vocal affect produced by youth and adults. However, emerging evidence suggests that the acoustic characteristics underlying both age groups' emotional expressions differ in ways that may make youth's emotional prosody more difficult to decode than adults' (Morningstar et al. 2017a). Specifically, youth actors' portrayals of socio-emotional expressions were less distinct in mean pitch than were those of adult actors: for example, though sadness was found to be expressed with significantly lower pitch than anger when produced by adult speakers, the same expressions spoken by youth were not significantly different in pitch. These findings suggest that youth's vocal representations of various emotional categories are less differentiated from one another than adults'. Given the importance of pitch cues to the recognition of vocal affect (Pell et al. 2009; Scherer 1996), youth's vocal expressions are likely to be less well-recognized than those of adults.

The presumed effect of speaker age on ER accuracy may also depend on the age of the listener. Due to the ongoing maturation of relevant brain areas during adolescence, it is likely that mid-adolescents may show greater vocal ER deficits relative to adults when hearing complex, harder-to-recognize stimuli, like youth-generated prosody. As such, though all listeners may have difficulty recognizing the intended affect in youth's emotional prosody, mid-adolescents may struggle even more to do so than adults. Given

adolescents' growing social networks (Larson and Richards 1991; Stanton-Salazar and Spina 2005), assessing youth's ability to decode such socially relevant cues can provide crucial information about the challenges they may face in navigating their social worlds.

Expression Type

Prior research on vocal ER has determined that certain emotions, such as anger and sadness, are better recognized than others, like happiness and disgust (e.g., Johnstone and Scherer 2000), suggesting expression-specific influences on vocal ER. Further, a growing body of work has established that the voice can convey important social information in attitudinal prosody (Banse and Scherer 1996; Cheang and Pell 2008; Juslin and Laukka 2003; Mitchell and Ross 2013). The current study thus examined listeners' recognition of both basic emotions (i.e., anger, disgust, fear, happiness, sadness) and social expressions denoting social hostility ("meanness") and affiliation ("friendliness"). Meanness has been defined as the commission or omission of acts that resulted or intended to hurt someone emotionally (Merten 1997), and may thus be similar to social aggression. Conversely, 'friendliness' can be conceptualized as social acceptance. These two expressions are theoretically and acoustically (Morningstar et al. 2017a) distinct from other basic emotions, such as anger and happiness, and may thus be perceived differently as well. Further, since understanding these expressions may require the listener to be able to infer others' intentions beyond their emotional state, the recognition of meanness and friendliness may require more advanced skills than the identification of basic emotions, and thus may be easier for adults to identify than for mid-adolescents. Since social expressions carry important communicative functions in interpersonal relationships, their inclusion in studies of nonverbal decoding is critical to our understanding of emotional communication processes.

Goals and Hypotheses of the Current Study

In the current investigation, we asked both adult and mid-adolescent (aged 13–15 years) listeners to identify the intended expression in vocal stimuli produced by both adult and youth (aged 10–15 years) actors in a previous investigation of age-related acoustic differences in emotional prosody (Morningstar et al. 2017a). Previous work has suggested that mid-adolescents should be adult-like in vocal ER skills (Brosigole and Weisman 1995; Zupan 2015); however, prior studies have typically assessed youth's recognition of a handful of basic emotions expressed by adult speakers. Given that our task included stimuli that may be more challenging to recognize, like youth-generated prosody and social expressions, we anticipated that adult listeners may outperform mid-adolescents overall.

We also examined whether the relative difference between adults' and mid-adolescents' performance varied as a function of stimulus-level variables. First, we hypothesized that listeners' performance would depend on the speaker's age. Specifically, we expected that adult actors' voices would be better-recognized by listeners than youth actors': adults' greater experience in conveying vocal affect, which may be reflected in their more differentiated emotional portrayals (Morningstar et al. 2017a), may facilitate listeners' recognition of the intended expression. In conjunction with this hypothesis, we expected that the effect of speaker age may interact with the age of the listener. Thus, although we predicted that all listeners would find youth voices more difficult to recognize than adult voices, mid-adolescent listeners may be even more disadvantaged than adult listeners when hearing expressions generated by other youth.

Second, we examined whether the type of expression would interact with listener age to predict vocal ER. Based on previous cross-sectional research (e.g., Brosgole and Weisman 1995; Zupan 2015), we expected that mid-adolescents would demonstrate accuracy equivalent to adults for some expressions (i.e., angry, happy, sad), but would not perform as well as adult listeners with more complex and later-acquired expressions, such as disgust (Widen and Russell 2013) or the social expressions of meanness and friendliness.

Lastly, previous work has documented the impact of gender on vocal ER. Specifically, meta-analyses have noted that females are more accurate decoders of vocal affect than are males (McClure 2000; Thompson and Voyer 2014). As well, several studies have reported that female speakers are better recognized than males (Belin, Fillion-Bilodeau, and Goselin 2008; Gallois and Callan 1986; Zuckerman et al. 1975). As such, we included both listener gender and speaker gender as control variables in our analysis.

Method

Participants

We recruited 55 mid-adolescent listeners and 89 adult listeners from a large Canadian city, from a database of families interested in research studies, as well as with flyers, advertisements in magazines and social media, and word of mouth. Adult participants' age was capped at 30 years old, given evidence that some adults demonstrate mild decline in emotion recognition ability beginning at 30 years of age (Mill et al. 2009). Eight of these participants were excluded from the sample prior to analyses: three adult listeners and one youth listener reported a dominant language other than English, and four youth listeners reported a diagnosis of speech disorder or hearing deficit.

The final sample consisted of 50 mid-adolescent listeners (58% female; age $M = 14.08$ years, $SD = 0.75$ years, range 13–15 years) and 86 adult listeners (59% female; age $M = 21.24$ years, $SD = 2.40$ years, range 18–30 years). Of these participants, 88.0% of youth listeners identified as Caucasian, 6.0% as of mixed ethnicity, and 6.0% as other ethnicities; 58.1% of adult listeners identified as Caucasian, 9.3% as of mixed ethnicity, and 32.6% as other ethnicities.

Stimuli

The expression recognition task comprised a subset of vocal recordings produced by adult and youth actors in a previous study examining age-related differences in emotional prosody (Morningstar et al. 2017a). The full set of recordings was produced by boys ($n = 7$, age $M = 12.51$ years old), girls ($n = 17$, age $M = 12.97$ years old), women ($n = 15$, age $M = 25.40$ years old), and men ($n = 15$, age $M = 33.20$ years old), who performed five neutral-content sentences in each of seven socio-emotional tones of voice: anger, happiness, sadness, fear, disgust, friendliness, and meanness.

The stimuli for the current study were chosen from the full set of recordings ($n = 3808$) based on ratings made by six independent raters (50% female; age $M = 22.33$ years old, $SD = 2.16$ years). The raters had no prior experience with the recordings used in the study, reported no hearing impairments, and spoke English as their dominant language. Raters were given a list of all recordings, along with the actors' intended expression. They then listened to each recording (as many times as needed), and scored them on recognizability

(“How recognizable is the emotion in this recording?”, on a four-point scale from 1 = ‘*not at all recognizable*’, to 4 = ‘*very much so*’), and authenticity (“How authentic is the emotion represented in this recording?”, on a four-point scale from 1 = ‘*not all authentic*’ to 4 = ‘*very much so*’; scales based on Banse and Scherer 1996) of the intended expression. Intraclass correlation coefficients (ICC) and their 95% confidence intervals (CI) were computed for raters’ judgments of recognizability and authenticity separately, based on a mean-rating ($k = 6$), consistency-based, two-way mixed model. Raters demonstrated acceptable reliability for both scales (Koo and Li 2016), with ICC = .80 (CI = .79–.81) for recognizability and ICC = .68 (CI = .66–.69) for authenticity.

For each recording, we computed a score for recognizability and authenticity by averaging all six raters’ ratings on each scale. Recordings with a score of < 2.5 (i.e., the mid-point) for either scale were discarded. Composite scores were created for the remaining recordings by averaging the recording’s recognizability and authenticity scores. The highest-scoring recordings for each actor group (e.g., boys, girls, men, women), expression, and sentence, were retained for inclusion in the study. In total, 35 recordings (7 expressions \times 5 sentences) were retained for each of the four actor groups (women, men, girls, and boys). This procedure selected a total of 140 recordings for the current study (see Table 1 for raters’ mean ratings of recognizability and authenticity for the selected recordings). All listeners heard recordings produced by both speaker age groups and genders.

Procedure

All procedures were approved by the institutional Research Ethics Board. Data were collected either in the lab or in quiet spaces of community settings, such as public libraries. All participants gave written consent/assent, and we obtained written parental consent for all youth. Participants then completed the expression recognition task. All 140 recordings were presented to listeners over sound-cancelling headphones, in a randomized order, using E-Prime stimulus presentation software. After hearing each recording twice, participants were asked to select the expression conveyed by the recording from a choice of 7 labels (anger, disgust, fear, friendliness, happiness, meanness, sadness). They were also asked to provide confidence ratings for their decision on a five-point scale (from 1 = “*I’m guessing*” to 5 = “*I’m certain*”, e.g., Pollak and Sinha 2002); these data will not be discussed here, but are available from the authors. Optional breaks were offered after each set of 50 recordings. Participants were then debriefed and compensated for their time.

Statistical Analyses

Each participant’s accuracy data was compiled into confusion matrices illustrating identification and error patterns for each speaker group (women, men, girls, and boys; see Table 2 for aggregated confusion matrices for all participants). From each matrix, we computed the unbiased hit rate¹ for each expression type (H_u ; Wagner 1993), which indexes accuracy correcting for participants’ response biases. A value of 1 for H_u would indicate perfect recognition (100% hit rate, without false alarms or false negatives), whereas a value of 0 would indicate no recognition of an expression (0% hit rate, with only false alarms and false negatives). To enable us to examine the impact of speaker

¹ H_u was computed as follows (Wagner 1993): (number of correct responses/number of times that stimuli category was presented) \times (number of correct responses/number of times that response was made).

Table 1 Raters’ judgments of selected recordings’ recognizability and authenticity

Expression	Recognizability		Authenticity	
	Youth speakers, <i>M (SD)</i>	Adult speakers, <i>M (SD)</i>	Youth speakers, <i>M (SD)</i>	Adult speakers, <i>M (SD)</i>
Anger	3.52 (0.75)	3.92 (0.28)	3.45 (0.62)	3.57 (0.62)
Disgust	3.05 (0.91)	3.37 (0.82)	3.25 (0.68)	3.38 (0.67)
Fear	3.50 (0.65)	3.52 (0.77)	3.55 (0.53)	3.40 (0.69)
Friendliness	3.52 (0.72)	3.83 (0.38)	3.23 (0.91)	3.50 (0.72)
Happiness	3.22 (0.80)	3.45 (0.75)	3.42 (0.59)	3.42 (0.65)
Meanness	3.27 (0.90)	3.58 (0.72)	3.52 (0.57)	3.42 (0.56)
Sadness	3.83 (0.49)	3.45 (0.70)	3.58 (0.62)	3.63 (0.52)

Values represent the mean (*M*) and standard deviation (*SD*) of 6 raters’ judgments of recognizability and authenticity for the recordings included in the expression recognition task. Both scales ranged from 1 (*not at all*) to 4 (*very much so*)

Table 2 Confusion matrices for youth and adults’ recognition of youth- and adult-generated expressions (% recognition)

Expression	Youth listeners							Adult listeners						
	A	D	F	Fr	H	M	S	A	D	F	Fr	H	M	S
Youth speakers														
Anger	47	14	18	3	2	11	6	62	6	17	2	0	8	5
Disgust	5	42	7	10	1	22	13	7	35	5	9	0	31	13
Fear	6	6	43	5	1	2	38	7	4	47	4	2	1	34
Friendliness	2	9	8	46	17	5	14	2	8	6	57	15	3	10
Happiness	1	6	2	49	34	3	5	1	5	2	54	31	3	4
Meanness	16	33	4	8	2	33	5	24	27	2	7	0	36	3
Sadness	0	1	17	2	1	1	78	0	2	11	2	1	0	85
Adult speakers														
Anger	79	3	5	0	0	12	0	83	3	4	0	0	9	0
Disgust	7	45	7	4	1	20	15	8	46	6	6	1	22	11
Fear	11	14	51	4	1	7	13	12	13	58	5	1	3	9
Friendliness	1	9	3	55	26	3	4	1	4	2	64	23	2	5
Happiness	4	6	5	36	41	5	3	4	4	2	42	43	3	3
Meanness	21	27	4	4	3	40	1	25	25	3	3	1	42	1
Sadness	2	9	15	7	1	2	64	2	8	10	8	0	1	72

Bold values highlight the accurate recognition of the intended expression type

A anger, D disgust, F fear, Fr friendliness, H happiness, M meanness, S sadness

characteristics (i.e., age and gender) on listeners’ accuracy, we derived H_u for each expression type conveyed by each group of speakers (boys, girls, men, and women) separately. This procedure resulted in 28 values of H_u for each listener (7 expressions \times 4 actor groups). A value of H_u cannot meaningfully be generated for listeners’ accuracy for a single recording; thus, we grouped speakers by their age group (youth vs. adults) instead of

looking at the impact of speakers' continuous age. Additionally, because H_u values are proportions, we followed Wagner's (1993) recommendation that they be arcsine transformed before use in analyses. Variables were screened for skewness and kurtosis, and distributions were sufficiently normal.

Results

A repeated-measures analysis of variance (ANOVA) was performed to examine the effects of listener age (between-subjects variable with 2 levels: youth vs. adult), speaker age (within-subjects variable with 2 levels: youth vs. adult), and expression type (within-subjects variable with 7 levels: anger, disgust, fear, friendliness, happiness, meanness, sadness) on listeners' accuracy (H_u). We also included listener gender (between-subjects variable with 2 levels: female vs. male) and speaker gender (within-subjects variable with 2 levels: female vs. male) as control variables in the model (see Appendix for full factorial model results). Greenhouse–Geisser corrections were used based on the results of Mauchly's test of sphericity (all p s < .001). Post-hoc pairwise comparisons with Šidák corrections and simple-effects tests were used to follow up on significant main effects and interactions, respectively.

There was a significant effect of listener age, $F(1, 132) = 14.49, p < .001, \eta^2 = .10$, such that adult listeners ($M = 1.17, SE = 0.02$) were more accurate than were mid-adolescent listeners ($M = 1.06, SE = 0.02$). A main effect of speaker age was also significant, $F(1, 132) = 140.71, p < .001, \eta^2 = .52$, such that recordings produced by adult speakers ($M = 1.22, SE = 0.02$) were better recognized than were those produced by youth speakers ($M = 1.01, SE = 0.02$). However, contrary to hypotheses, the two-way interaction between listener and speaker age was not significant, $F(1, 132) = 0.23, p = .64, \eta^2 < .01$.

We then examined whether the type of expression being conveyed was associated with listeners' accuracy. First, there was a main effect of expression type, $F(5.28, 696.76) = 130.89, p < .001, \eta^2 = .50$. Post-hoc pairwise comparisons with Šidák corrections suggested that sadness ($M = 1.51, SE = 0.02$) and anger ($M = 1.45, SE = 0.03$) were best recognized (and not different from one another, $p > .05$), followed by fear ($M = 1.14, SE = 0.03$) and friendliness ($M = 1.06, SE = 0.02$; the latter two were not different from one another, $p > .05$), happiness ($M = 0.97, SE = 0.03$; not different from friendliness, $p > .05$), and disgust ($M = 0.87, SE = 0.02$; not different from happiness, $p > .05$) and meanness ($M = 0.83, SE = 0.03$; the latter two did not differ from one another, $p > .05$). Unless otherwise noted above, all pairwise differences between expression types were significant, all p s < .05.

Additionally, there were significant two-way interactions between both listener age and expression type, and speaker age and expression type, which were qualified by a three-way interaction between listener age, speaker age, and expression type, $F(5.25, 696.70) = 3.55, p < .01, \eta^2 = .03$ (see Fig. 1). We unpacked this interaction by conducting separate ANOVAs on the effect of listener age within each level combination of speaker age and expression type. These tests revealed that adults were more accurate than youth when hearing adult portrayals of fear, $F(1, 132) = 10.14, p < .01, \eta^2 = .07$, and sadness, $F(1, 132) = 12.21, p = .001, \eta^2 = .09$, and when hearing youth portrayals of anger, $F(1, 132) = 15.63, p < .001, \eta^2 = .11$, fear, $F(1, 132) = 4.35, p < .05, \eta^2 = .03$, friendliness, $F(1, 132) = 16.11, p < .001, \eta^2 = .01$, and sadness, $F(1, 132) = 11.64,$

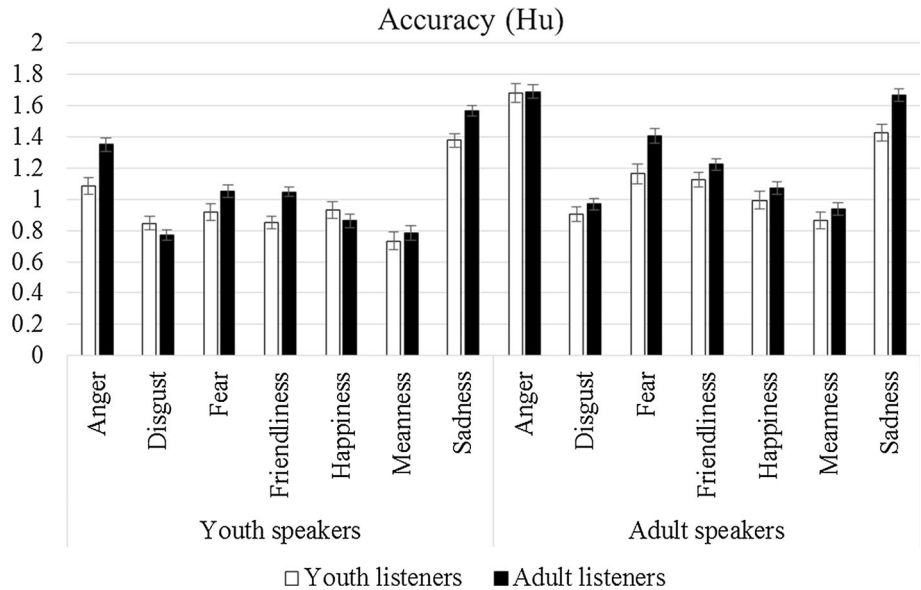


Fig. 1 Three-way interaction between speaker age, listener age, and expression on listener accuracy (H_u ; unbiased hit rate, arcsine-transformed). Error bars represent the standard error of the mean

$p = .001$, $\eta^2 = .08$. Mid-adolescent and adult listeners were equivalent in performance for adult portrayals of anger, disgust, friendliness, happiness, and meanness, and for youth portrayals of disgust, happiness, and meanness (all $ps > .05$).

Inspection of the confusion matrices for mid-adolescent and adult listeners' recognition of youth and adult voices revealed common error patterns. In descriptive terms, fear and sadness were often mistaken for one another, especially when the exemplars were produced by youth speakers. Happiness and friendliness were also often confused asymmetrically, such that happiness tended to be mistaken for friendliness more often than the opposite; this pattern seemed especially pronounced when listeners heard youth's portrayals of these expressions.

Discussion

The current study examined whether listener age interacted with the stimulus-level properties of speaker age and type of expression to predict youth's and adults' recognition of seven vocal socio-emotional expressions. As hypothesized, adult listeners were more skilled in recognizing emotional prosody than were 13- to 15-year-old mid-adolescents. Though prior work has suggested that youth achieve adult-like accuracy in vocal ER tasks around the age of 12 (Brosgole and Weisman 1995; Zupan 2015), our findings indicate that these skills continue to develop in mid-adolescence and beyond. Such results are consistent with neuroimaging evidence that brain areas involved in socio-emotional processing develop dramatically during adolescence (Blakemore and Choudhury 2006; Kilford et al. 2016), changes that presumably support increased accuracy in vocal ER. Our study aimed to determine whether these age-related improvements in ER varied based on factors inherent to the stimuli, namely the age of the speaker and the type of expression conveyed.

The difference in performance between adult and youth listeners depended both on the expression they were hearing, and who was speaking.

Speaker Age

As predicted, youth's vocal expressions were overall more difficult to identify than those of adults. Adults may be able to better communicate emotional information than their younger counterparts, due to both increased experience with age, and developmental changes in the laryngeal tract (e.g., Fitch and Giedd 1999) that may facilitate adults' greater vocal differentiation between expressions (Morningstar et al. 2017a). Although adults' greater encoding skill relative to youth may be due to the maturation of emotional skills with age, we cannot rule out the possibility that this effect reflects a difference in acting skills, and that adult speakers were simply more skilled actors than youth. However, youth speakers were also actors, whose portrayals of socio-emotional expressions may have been modeled after their adult acting teachers' expressions. Youth actors' representations of various expressions may therefore be more similar to adults' than recordings made by non-actor youth; in such a case, our design may be underestimating the difference between listeners' recognition of youth and adult speakers' emotional outputs. Though we cannot disentangle the effect of age from that of acting experience in the current design, the greater experience of adult actors compared to youth actors reflects the pattern in the general population: adults, in general, have had more practice communicating affect than youth, and listeners are better able to interpret their emotional prosody.

We hypothesized that, though all listeners would struggle to identify youth voices, mid-adolescent listeners would be more disadvantaged than adults with these difficult stimuli. A three-way interaction between speaker age, listener age, and expression revealed that the performance difference between both listener groups was indeed greater with youth voices. When hearing adult-generated expressions, mid-adolescent listeners were equivalent in accuracy to adults for most expressions under investigation (i.e., anger, disgust, friendliness, happiness, and meanness), and were less accurate only when hearing fear and sadness. However, when listening to youth actors' voices, mid-adolescents identified most expressions (i.e., anger, fear, friendliness, and sadness) less accurately than adults did. Thus, though mid-adolescents generally achieved adult-like recognition for adult-generated voices, our findings suggest that the development of their ER skills may not be fully complete with the harder-to-identify youth voices. In fact, ongoing development of ER skills may primarily be evident with peer-aged cues at this age.

It is likely that mid-adolescents' poorer recognition of most youth-produced expressions compared to adults is due to the incomplete development of both the decoders' ER ability and of the encoders' communication skill. Though these effects are difficult to disentangle, our data indicate that the developmental level of the encoder may matter more for vocal ER than the developmental level of the decoder. Indeed, the magnitude of the association between speaker age and recognition accuracy was more than 5 times larger than that of the relation between listener age and accuracy. This pattern was also reflected in decoders' errors, which were similar in nature across listener age groups, but more pronounced with youth stimuli than adult versions. Together, these results highlight the importance of speaker-level variables for listeners' vocal ER accuracy. Little is known about factors that may systematically shape the ability to express paralinguistic cues: given the associations between encoding skill and positive social outcomes in childhood (Boyatzis and Satya-prasad 1994) and adulthood (Riggio and Friedman 1986), further research is needed to

elucidate experiences and characteristics that contribute to the ability to be perceived accurately.

Expression Type

Though we expected adults' superior accuracy compared to youth listeners to be evidenced with poorly-recognized and complex expressions, adult listeners outperformed their younger counterparts primarily with well-identified expressions, such as anger, sadness, and fear. Beyond the possible non-specific impact of increased neural maturation and exposure to emotional cues, it is unclear why adults were more accurate than mid-adolescents with these particular expressions. Moreover, previous studies have suggested that youth should be adult-like in recognizing expressions of sadness and fear by early adolescence (Brosigole and Weisman 1995; Zupan 2015). We offer two possible reasons for our divergent findings. First, the adolescent sample included in Brosigole and Weisman's (1995) study ranged from 13 to 17 years of age. Our sample included only the lower end of that age range, suggesting that adult-like abilities in the recognition of fear and sadness may emerge at a later developmental stage. Second, our ER task may have been more difficult than that of both previous investigations, which reported generally high accuracy rates overall (with potential ceiling effects in Brosigole and Weisman 1995). Given that our ER task included youth voices and seven expressions rather than the typical three or four, the complexity of the task may have suppressed youth's accuracy in comparison to prior studies. It will be important to attempt to replicate these results, and to explore possible mechanisms by which certain vocal expressions may become easier to identify with age.

In contrast, adults' superior accuracy compared to youth did not extend to expressions that are more difficult to identify, such as happiness and disgust (Johnstone and Scherer 2000; Scherer 2003), regardless of speaker age. Disgust, meanness, and happiness may be hard to convey using vocal prosody. Indeed, happiness is often found to be easier to identify in the face than in speech prosody (e.g., Hawk et al. 2009), and it has been suggested that disgust may be more easily communicated by vocal bursts (e.g., "ugh!"); Banse and Scherer 1996). Similarly, it is possible that meanness is conveyed more accurately via postures and gestures, such as the mean glares and faces that denote social aggression in teenage girls (LaFrance 2002; Simmons 2002; Underwood 2004), than with verbal content. Hence, adults' increased maturation and experience with emotion cues may not be sufficient to confer a processing advantage with expressions that are difficult to both communicate and identify using acoustic cues.

Strengths and Limitations

The current study explicitly contrasted mid-adolescent and adult listeners' ability to identify expressions conveyed by both adult and youth speakers. This design permitted us to examine the influence of speaker characteristics on vocal ER, and to evaluate youth's recognition of vocal cues produced by two important groups in their social environments. Youth's interpretation of other adolescents' emotional outputs is particularly important for success in social interactions during this developmental stage. Our findings highlight a potential challenge in youth's social lives: the vocal prosody of their peers, stimuli that are crucial to their interpersonal interactions, is hard to decode. Describing the difficulties inherent to navigating their socio-emotional worlds helps inform our understanding of peer interactions in adolescence.

As well, this study extended earlier work by examining listeners' ability to identify both basic emotions and social expressions. Though vocal expressions of friendliness were recognized at similar rates as happiness, cues of meanness were difficult for listeners to identify. Indeed, meanness yielded one of the lowest accuracy rates of all expressions, and was significantly harder to recognize than anger. Our results suggest that further research is needed to uncover the types of social cues that can reliably be communicated by the voice. Further, examining listeners' recognition of a broad array of expressions allowed us to qualify adults' overall ER advantage over mid-adolescents, by noting expressions for which youth were equivalent in accuracy to adults. Our findings support the inclusion of a variety of expressions in ER tasks, in order to provide a more complete picture of listeners' vocal recognition skills.

Limitations must also be noted. First, the recordings selected for the expression recognition task were chosen based on ratings of authenticity and recognizability made by adults. We did not think it was feasible to have youth complete this task, which involved rating over 3000 recordings; however, it is possible that younger listeners may have chosen different recordings as prototypical representations of the relevant expressions. If so, mid-adolescents may have performed better with recordings they selected, minimizing the effect of listener age on accuracy. However, a large effect of listener age on vocal ER was also noted in a study in which youth-generated stimuli were selected based on both youth and adult listeners' recognition rates, instead of adult judges' ratings (Morningstar et al., 2017b). As such, the effect of listener age is probably robust regardless of the selection method used to choose task recordings. Nonetheless, future research should investigate whether adult and youth raters differ in their judgments of the quality and representativeness of emotional exemplars.

Second, our task design did not allow participants to select a "none of the above" option in response to the recordings they heard. It has been suggested that the omission of such a label can lead to artificial agreement in forced-choice paradigms (e.g., Frank and Stennett 2001). It is possible that including this type of response option may have led to decreased accuracy rates overall or that different error patterns would have emerged, although it seems less likely that adding this response label would have affected youth and adult listeners differently or influenced the main effects noted in the current study. Nonetheless, free-labelling paradigms may be valuable alternatives to forced-choice tasks to provide a more nuanced understanding of listeners' interpretation of emotional prosody, particularly for expressions beyond basic emotions.

Third, the recordings we used as stimuli were prototypical representations of various expressions generated by actors. It is possible that the patterns found in the current study may have been different with naturalistic recordings of untrained speakers, though select studies have found that actors and non-actors elicited similar recognition patterns by listeners (Jürgens et al. 2015; Spackman, Brown, and Otto 2009). Future research should include naturalistic representations of emotion to supplement lab-based findings on vocal ER development.

Conclusion

Results of this investigation suggest that age-related changes in the recognition of vocal emotional cues can vary as a function of stimuli-level properties. Consistent with neurological evidence demonstrating protracted development of brain areas involved in socio-

emotional processing in adolescence (Kilford et al. 2016), adult listeners in our study were more accurate in decoding vocal expressions than mid-adolescent listeners. However, this effect varied depending on both the age of the speaker producing the emotional prosody, and the type of expression being conveyed.

We found that adult speakers were better recognized than were youth speakers, highlighting the potential impact of speaker characteristics on listeners' accuracy. Future work should systematically vary characteristics of the speaker along theoretically meaningful dimensions, to provide insight into mechanisms that contribute to the ability to express socio-emotional information vocally. Moreover, though mid-adolescent listeners demonstrated adult-like recognition of adult vocal cues, they were less accurate at identifying most youth-generated expressions than adult listeners were. Taken together, our findings suggest that the task of interpreting peers' emotional outputs may be more difficult for youth than for adults. Errors in these interpretations may present a unique social challenge for adolescents, and may be a fruitful target of social skills training for youth who struggle socially.

Funding This work was supported by the Social Sciences and Humanities Research Council of Canada.

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: Full Factorial ANOVA Model Results

Vocal cues	<i>df</i>	<i>F</i>	η^2	<i>p</i>
Listener age	(1, 132)	14.49	.10	< .001
Listener gender	(1, 132)	10.42	.07	< .01
Speaker age	(1, 132)	140.71	.52	< .001
Speaker gender	(1, 132)	4.84	.04	< .05
Expression	(5.28, 696.76)	130.89	.50	< .001
Speaker age × listener age	(1, 132)	0.23	< .01	.64
Speaker gender × listener gender	(1, 132)	0.01	< .001	.93
Listener age × listener gender	(1, 129)	0.09	.001	.76
Listener age × speaker gender	(1, 132)	4.19	.03	.04
Speaker age × listener gender	(1, 132)	0.12	.001	.73
Speaker age × speaker gender	(1, 132)	2.03	.02	.16
Expression × listener age	(5.28, 696.96)	3.54	.03	< .01
Expression × listener gender	(5.28, 696.96)	0.54	< .01	.76

Vocal cues	<i>df</i>	<i>F</i>	η^2	<i>p</i>
Expression × speaker age	(5.25, 692.39)	12.02	.08	< .001
Expression × speaker gender	(5.34, 705.42)	27.49	.17	< .001
Expression × listener age × listener gender	(5.28, 681.12)	1.93	.01	.08
Speaker age × listener age × listener gender	(1, 129)	0.18	.001	.67
Speaker gender × listener age × listener gender	(1, 129)	0.93	.01	.34
Speaker age × speaker gender × listener age	(1, 132)	1.65	.01	.20
Speaker age × speaker gender × listener gender	(1, 132)	0.80	.01	.37
Speaker age × expression × listener age	(5.25, 696.96)	3.55	.03	< .01
Speaker age × expression × listener gender	(5.25, 696.96)	1.68	.01	.13
Speaker gender × expression × listener age	(5.34, 696.96)	1.53	.01	.17
Speaker gender × expression × listener gender	(5.34, 696.96)	1.60	.01	.15
Speaker age × speaker gender × expression	(5.08, 670.26)	11.04	.08	< .001
Speaker age × speaker gender × listener age × listener gender	(1, 129)	3.16	.02	.08
Speaker age × expression × listener age × listener gender	(5.25, 681.12)	0.25	< .01	.95
Speaker gender × expression × listener age × listener gender	(5.34, 681.12)	0.95	.01	.45
Speaker age × speaker gender × expression × listener age	(5.08, 696.96)	1.19	.01	.31
Speaker age × speaker gender × expression × listener gender	(5.08, 696.96)	1.37	.01	.23
Speaker age × speaker gender × expression × listener age × listener gender	(5.08, 681.12)	1.01	.01	.41

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. <https://doi.org/10.1111/bjdp.12097>.
- Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*, 614–636. <https://doi.org/10.1037/0022-3514.70.3.614>.
- Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal affective voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behaviour Research Methods*, *40*, 531–539. <https://doi.org/10.3758/BRM.40.2.531>.
- Blakemore, S.-J., & Choudhury, S. (2006). Development of the adolescent brain: Implications for executive function and social cognition. *Journal of Child Psychology and Psychiatry*, *47*, 296–312. <https://doi.org/10.1111/j.1469-7610.2006.01611.x>.
- Blasi, A., Mercure, E., Lloyd-Fox, S., Thomson, A., Brammer, M., Sauter, D., et al. (2011). Early specialization for voice and emotion processing in the infant brain. *Current Biology*, *21*, 1220–1224. <https://doi.org/10.1016/j.cub.2011.06.009>.

- Boyatzis, C. J., & Satyaprasad, C. (1994). Children's facial and gestural decoding and encoding: Relations between skills and with popularity. *Journal of Nonverbal Behavior*, 18, 37–55. <https://doi.org/10.1007/BF02169078>.
- Brosigole, L., & Weisman, J. (1995). Mood recognition across the ages. *International Journal of Neuroscience*, 82, 169–189. <https://doi.org/10.3109/00207459508999800>.
- Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.
- Cheang, H. S., & Pell, M. D. (2008). The sound of sarcasm. *Speech Communication*, 50, 366–381. <https://doi.org/10.1016/j.specom.2007.11.003>.
- Chronaki, G., Hadwin, J. A., Garner, M., Mauraige, P., & Sonuga-Barke, E. J. S. (2015). The development of emotion recognition from facial expressions and non-linguistic vocalizations during childhood. *British Journal of Developmental Psychology*, 33, 218–236. <https://doi.org/10.1111/bjdp.12075>.
- 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.
- Frank, M. G., & Stennett, J. (2001). The forced-choice paradigm and the perception of facial expressions of emotion. *Journal of Personality and Social Psychology*, 80, 75–85.
- Gallois, C., & Callan, V. J. (1986). Decoding emotional messages: Influence of ethnicity, sex, message type, and channel. *Journal of Personality and Social Psychology*, 51, 755–762. <https://doi.org/10.1037/0022-3514.51.4.755>.
- Grossmann, T. (2010). The development of emotion perception in face and voice during infancy. *Restorative Neurology and Neuroscience*, 28, 219–236. <https://doi.org/10.3233/RNN-2010-0499>.
- Grossmann, T., Oberecker, R., Koch, S. P., & Friederici, A. D. (2010). The developmental origins of voice processing in the brain. *Neuron*, 65, 852–858. <https://doi.org/10.1016/j.neuron.2010.03.001>.
- Hawk, S. T., van Kleef, G. A., Fischer, A. H., & van der Schalk, J. (2009). “Worth a thousand words”: Absolute and relative decoding of nonlinguistic affect vocalizations. *Emotion*, 9, 293–305. <https://doi.org/10.1037/a0015178>.
- 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. ISBN-10: 1609180445.
- Jürgens, R., Grass, A., Drolet, M., & Fischer, J. (2015). Effect of acting experience on emotion expression and recognition in voice: Non-actors provide better stimuli than expected. *Journal of Nonverbal Behavior*, 39, 195–214. <https://doi.org/10.1007/s10919-015-0209-5>.
- Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, 129, 770–814. <https://doi.org/10.1037/0033-2909.129.5.770>.
- Kilford, E. J., Garrett, E., & Blakemore, S. J. (2016). The development of social cognition in adolescence: An integrated perspective. *Neuroscience and Biobehavioral Reviews*, 70, 106–120.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15, 155–163.
- LaFrance, M. (2002). Smile boycotts and other body politics. *Feminism & Psychology*, 12, 319–323. <https://doi.org/10.1177/0959353502012003003>.
- Larson, R., & Richards, M. H. (1991). Daily companionship in late childhood and early adolescence: Changing developmental contexts. *Child Development*, 62, 284–300. <https://doi.org/10.1111/j.1467-8624.1991.tb01531.x>.
- McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children and adolescents. *Psychological Bulletin*, 126, 424–453. <https://doi.org/10.1037/0033-2909.126.3.424>.
- Merten, D. E. (1997). The meaning of meanness: Popularity, competition, and conflict among junior high school girls. *Sociology of Education*, 70, 175–191.
- Mill, A., Allik, J., Realo, A., & Valk, R. (2009). Age-related differences in emotion recognition ability: A cross-sectional study. *Emotion*, 9, 619–630. <https://doi.org/10.1037/a0016562>.
- 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. <https://doi.org/10.1016/j.neubiorev.2013.01.027>.
- Morningstar, M., Dirks, M. A., & Huang, S. (2017a). Vocal cues underlying youth and adult portrayals of socio-emotional expressions. *Journal of Nonverbal Behavior*, 41, 155–183. <https://doi.org/10.1007/s10919-017-0250-7>.
- Morningstar, M., Dirks, M. A., Rappaport, B. I., Pine, D. S., & Nelson, E. E. (2017b). Associations between anxious and depressive symptoms and the recognition of vocal socio-emotional expressions in youth. *Journal of Clinical Child & Adolescent Psychology*. <https://doi.org/10.1080/15374416.2017.1350963>.

- Nelson, E. E., Leibenluft, E., McClure, E. B., & Pine, D. S. (2005). The social re-orientation of adolescence: A neuroscience perspective on the process and its relation to psychopathology. *Psychological Medicine*, 35, 163–174. <https://doi.org/10.1017/S0033291704003915>.
- Nowicki, S., & Duke, M. P. (2001). Individual differences in the nonverbal communication of affect: The diagnostic analysis of nonverbal accuracy scale. *Journal of Nonverbal Behavior*, 18, 9–35.
- Pell, M. D., Paulmann, S., Dara, C., Alasserri, A., & Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics*, 37, 417–435. <https://doi.org/10.1016/j.wocn.2009.07.005>.
- Pollak, S. D., & Sinha, P. (2002). Effects of early experience on children's recognition of facial displays of emotion. *Developmental Psychology*, 38, 784–791. <https://doi.org/10.1037/0012-1649.38.5.784>.
- Riggio, R. E., & Friedman, H. S. (1986). Impression formation: The role of expressive behavior. *Journal of Personality and Social Psychology*, 50, 421–427. <https://doi.org/10.1037//0022-3514.50.2.421>.
- Rothman, A. D., & Nowicki, S. (2004). A measure of the ability to identify emotion in children's tone of voice. *Journal of Nonverbal Behavior*, 28, 67–92.
- Sauter, D. A., Panattoni, C., & Happé, F. (2012). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology*, 31, 97–113. <https://doi.org/10.1111/j.2044-835X.2012.02081.x>.
- 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. [https://doi.org/10.1016/S0167-6393\(02\)00084-5](https://doi.org/10.1016/S0167-6393(02)00084-5).
- Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion*, 15, 123–148. <https://doi.org/10.1007/BF00995674>.
- Simmons, R. (2002). *Odd girl out: The hidden culture of aggression in girls*. New York: Harcourt. ISBN 054735102X.
- 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. <https://doi.org/10.1080/02699930802536268>.
- Stanton-Salazar, R. D., & Spina, S. U. (2005). Adolescent peer networks as a context for social and emotional support. *Youth & Society*, 36, 379–417. <https://doi.org/10.1177/0044118X04267814>.
- Thompson, A. E., & Voyer, D. (2014). Sex differences in the ability to recognise non-verbal displays of emotion: A meta-analysis. *Cognition and Emotion*, 28, 1164–1195. <https://doi.org/10.1080/02699931.2013.875889>.
- 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. <https://doi.org/10.1177/0959353504044637>.
- Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior*, 17, 3–28. <https://doi.org/10.1007/BF00987006>.
- Widen, S. C., & Russell, J. A. (2013). Children's recognition of disgust in others. *Psychological Bulletin*, 139, 271–299. <https://doi.org/10.1037/a0031640>.
- Zuckerman, M., Lipets, M. S., Hall Koivumaki, J., & Rosenthal, R. (1975). Encoding and decoding non-verbal cues of emotion. *Journal of Personality and Social Psychology*, 32, 1068–1076. <https://doi.org/10.1037/0022-3514.32.6.1068>.
- 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. Retrieved from: <http://files.aiscience.org/journal/article/html/70320049.html>.