The Pennsylvania State University

The Graduate School

College of the Liberal Arts

CHILDREN RELY ON SPEAKER FAMILIARITY, BUT NOT ACOUSTIC PROPERTIES, TO DETERMINE EMOTIONAL INTENSITY

A Thesis in

Psychology

by

Tawni Blaze Stoop

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

December 2018

The thesis of Tawni Blaze Stoop was reviewed and approved* by the following:

Pamela M. Cole
Professor of Psychology
Thesis Advisor

Ginger A. Moore
Associate Professor of Psychology

Rick O. Gilmore
Associate Professor of Psychology

Melvin M. Mark
Professor of Psychology
Head of the Department of Psychology

*Signature are on file in the Graduate School

Abstract

This thesis examined the relations between acoustic properties that determine affective prosody, namely mean fundamental frequency (F0), F0 standard deviation, and speech rate, speaker familiarity, and children's perceptions of emotional intensity for angry, happy, and sad prosodies. We hypothesized that the acoustic properties would significantly predict children's ratings of emotional intensity. Further, we hypothesized that children would perceive their own mothers as being more intensely emotional than unfamiliar mothers, and that this speaker familiarity effect would account for a significant amount of variance in children emotional intensity ratings over and above the acoustic properties. This thesis employed a series of simultaneous regression analyses, a MANOVA model, and a series of hierarchical regressions. The results partially supported the hypotheses. Children rated their own mothers as more intensely emotional compared to the unfamiliar mothers, and this relationship depended on the emotion being conveyed. Speaker familiarity accounted for a significant amount of variance in children's emotional intensity ratings over and above the acoustic properties only for the angry prosody recordings. However, the acoustic properties of F0, F0 standard deviation, and speech rate were not associated with children's emotional intensity ratings for any of the affective prosodies. Interpretation of these findings, study limitations, and future directions are discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Institute of Mental Health under Grant No. 1R21MH104547-01A1 awarded to Pamela M. Cole, Ph.D. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Institute of Mental Health or the National Institutes of Health.

Chapter 1. INTRODUCTION

Consider a mother asking her child, “What are you doing?” A child might interpret the question differently if that mother asks in a light and joking tone, compared to a sterner, annoyed tone of voice. This difference could mean that the child continues the activity joyfully, or abruptly stops. What if that mother asks that same question angrily, but in a quieter, exasperated way as opposed to yelling loudly? A child may anticipate different outcomes depending on the intensity with which the mother conveys her emotion. As this example demonstrates, the same words convey different meaning depending upon the speaker’s affective prosody, the non-linguistic acoustic properties comprising emotional information, and emotional intensity. Now imagine that another woman, someone unfamiliar to the child, asks the same question. Again, how the child will interpret the meaning and significance of the question also depends on the child’s familiarity with that speaker. Mild joy or anger in the mother’s voice may be more meaningful to a child than the same emotions expressed by an unfamiliar woman. The social information conveyed through speech is complex, and the ability to read nonverbal information is critical to effective social interaction (Adolphs, 2002). Moreover, we know that the ability to decode emotional cues is one important component of the development of socioemotional competence in childhood (Halberstadt, Denham, & Dunsmore, 2001; McElwain, Halberstadt, & Volling, 2007; Saarni, 1999). Children’s understanding of others’ emotions involves more than just comprehending the words that they say. It includes *how* those words are spoken and who is speaking. The current study examined the relations between children’s perceptions of the intensity of speakers’ emotions and three features of speech stimuli: the specific emotion intended by the speaker, the acoustic properties, or affective prosodic information, of different specific emotions, and speaker familiarity.

Affective Prosody

The term affective prosody refers to paralinguistic information that carries the emotion information conveyed in speech. It influences how individuals process and recognize emotional semantic content (Nygaard & Queen, 2008). Importantly, affective prosody in speech adds information that can shape listeners' expectations beyond what is available in semantic cues alone (Paulmann, Titone, & Pell, 2012). Sensitivity to variations in affective prosody develops early in life, such that preverbal infants distinguish between prosodies (Fernald, 1993; Walker-Andrews & Grolnick, 1983). For example, 5-month-olds show facial expressions that affectively match approving and disapproving vocalizations (Fernald, 1993). As they enter the second year, infants modify their behavior based on variations in adults' affective prosody (Mumme, Fernald, & Herrera, 1996; Repacholi & Meltzoff, 2007; Vaish & Striano, 2004). For example, 18-month-olds who have learned to perform a simple task stop completing the task after an unfamiliar adult speaks angrily at the research assistant teaching the task to the toddlers (Repacholi & Meltzoff, 2007). In the preschool years, two eye-tracking studies suggest that 4- and 5-year-olds pre-emptively look at an object (broken vs. intact toy) that matches the affective prosody of semantically neutral instructions before the target object is identified (Berman, Chambers, & Graham, 2010; Berman, Graham, & Chambers, 2013).

Although these studies provide evidence of the ability to discriminate among some affective prosodies in the first five years of life, other studies present mixed results regarding whether children under the age of 7 years can identify emotion in the voice relative to older children. For example, children between the ages of 3 through 6 years old label emotions at or below chance rates of accuracy in non-emotional speech (Aguert, Laval, Lacroix, Gil, & Le Bigot, 2013; Nelson & Russell, 2011). Additional evidence using congruent and incongruent verbal-nonverbal stimuli indicates that younger children rely on semantic cues more than older children do (Friend, 2000; Friend & Bryant, 2000; Morton & Trehub, 2001). However, other studies suggest that 3- to 6-year-olds accurately identify emotions based on affective prosody with unintelligible (Baltaxe, 1991; Morton, Trehub, & Zelazo, 2003) or semantically non-emotional speech (Sauter, Panattoni, & Happé, 2013). One possible explanation is that in experimental situations, young children may rely on semantic information when presented with difficult stimuli that include contradictory semantic and prosodic information (Friend, 2000; Friend, 2003; Friend & Bryant, 2000; Morton & Trehub, 2001), but with more familiar, less complicated stimuli they may understand and use affective prosody accurately (Baltaxe, 1991; Berman et al., 2010). In fact, children's accuracy identifying prosody in experimental stimuli improves when children are primed to focus on prosody (Waxer & Morton, 2011), further highlighting that they are sensitive to prosody.

Research consistently agrees, however, that children's accuracy using affective prosody to label emotions increases as they age (Aguert et al., 2013; Friend, 2000; Morton & Trehub, 2001; Nelson & Russell, 2011; Rothman & Nowicki, 2004; Sauter et al., 2013). Typically developing school age children perform as well as adults in emotion labeling based on affective prosody (Van Lancker, Cornelius, & Kreiman, 1989). By age 9, children perform at 80% accuracy in labeling emotion based on prosody even if they believe they are relying on neutral contextual information (Aguert et al., 2013). Additionally, approximately half of 10-year-old participants primarily use affective prosody instead of semantic information to identify emotions in conflicting semantic and prosodic contexts (Morton & Trehub, 2001). Taken together, these studies suggest that very young children distinguish among basic emotions conveyed through prosody (e.g., negative versus positive, anger versus happy) and become more skilled at handling more complex stimuli, including integrating semantic and paralinguistic information even when they are incongruent, as they age. By early school age, 7- and 8-year-olds accurately identify emotion in the voice, particularly if presented in ways that are congruent or semantically neutral.

Emotional Intensity

Interpreting vocal emotion involves more than just emotion labeling. The majority of research on children's accuracy at reading nonverbal emotion cues, however, has focused on labeling emotions (Friend, 2000; Morton & Trehub, 2001; Sauter et al., 2013). Although it is important that children accurately label emotions, they must also interpret other features of the emotions they hear in the voices around them. Day-to-day social interactions involve dynamic and nuanced information. In many situations, children's perceptions of the intensity of emotion in the voice may be more meaningful than accurate recognition alone.

There is not a large volume of research on children's perception of the intensity of emotion cues. The facial emotion literature suggests that typically developing children struggle to identify lower-intensity emotion expressions (Herba, Landau, Russell, Ecker, & Phillips, 2006). On the other hand, children from families with documented physical abuse require fewer facial cues to identify anger, such that they accurately identify anger at a lower intensity compared to their peers (Pollak & Sinha, 2002). For these children, perceiving low-level anger may help them avoid unwanted consequences. Much like how facial expressions vary in emotional intensity, vocal expressions convey ranges of emotional intensity as well; unlike faces, we can interpret vocal cues from a distance, without seeing a person. Children's emotion labeling accuracy does not differ for high- versus low-intensity vocal expressions (Zupan, 2015), however it is still unclear how children perceive differences in vocal emotion intensity at all and the extent to which familiarity with a speaker influences children's perceptions of emotion intensity.

Acoustic Properties

Young children's sensitivity to affective prosody raises the question of their sensitivity to acoustic cues when they determine a speaker's emotion or emotional intensity. Affective prosody comprises variations in acoustic properties by which listeners distinguish between emotions (Banse & Scherer, 1996; Williams & Stevens, 1972). For adults, specific emotion labels are associated with distinct acoustic profiles, notably including variations in mean fundamental frequency (F0; pitch), variability of F0 (the spread of pitch values), and duration of voiced periods (speech rate typically measured in words per second), among other properties (Banse & Scherer, 1996; Banziger & Scherer, 2005; Breitenstein, Van Lankcer, & Daum, 2001; Scherer, Koivumaki, & Rosenthal, 1972; Streeter, Macdonald, Apple, Krauss, & Galotti, 1983). In particular, higher mean F0, or average pitch (Bachorowski & Owren, 2003; Breitenstein et al., 2001; Pell, Paulmann, Dara, Alasseri, & Kotz 2009), as well as F0 variability, or wider range of pitch values used (Banse & Scherer, 1996; Sobin & Alpert, 1999), and faster speech rate (Breitenstein et al., 2001) are associated with anger and happiness, especially when compared to sadness (Pell et al., 2009). Anger also has a higher mean F0, wider F0 variability, and slower speech rate than happiness (Pell et al., 2009). Though vocal emotions exist on a dimensional scale of intensity much like the dimensional nature of acoustic properties of pitch and speech rate, little research has actually focused on their relation to one another. Studies using adult participants demonstrated that fluctuations in mean F0 and F0 variability were associated with ratings of the speaker's intensity of stress (Juslin & Laukka, 2001; Streeter et al., 1983).

Given the evidence that children, like adults, are able to decode affective prosody, it is possible that these acoustic properties influence children's vocal emotion processing, including how they rate their intensity. For example, child-directed speech typically has higher F0, larger F0 variability, and moderate speech rate compared to adult-directed speech (Fernald & Mazzie, 1991), similar acoustic characteristics to a happy prosody. Infants demonstrate a preference for this type of child-directed speech compared to adult-directed speech (Schachner & Hannon, 2011). Surprisingly, this preference for child-directed speech seems to be more about the similarity to happy prosody, based on one study that showed no difference in preference when prosodies for child-directed and adult-directed speech were held constant (Singh, Morgan, & Best, 2002). This suggests that infants are focusing on prosodic cues like high F0 to drive preferences. However, little research has actually examined these acoustic properties in relation to children's understanding of emotion in the voice. One study suggested that 4- and 5-year-old children successfully used exaggerated pitch (F0) contours to determine if a puppet was happy or sad (Quam & Swingley, 2012). Additionally, the mean F0 has been associated with the speed and accuracy of emotion recognition of joy and anger in children aged 7-17 (Dmitrieva, Gel'man, Zaitseva, & Orlov, 2008). Speech rate did not have a significant relation with emotion recognition in these children (Dmitrieva et al., 2008), however this study did not examine recognition of sad stimuli which is the emotion most related to speech rate (Williams & Stevens, 1972). This limited work unfortunately focuses on emotion labeling. There is a lack of research examining how children's perceptions of emotional intensity in vocal cues relate to these acoustic properties. Further research is needed to more thoroughly understand how these acoustic properties influence children's interpretation of vocal emotion beyond emotion identification.

The current study fills this gap by addressing how those acoustic properties of F0 (pitch) and speech rate relate to their ratings of emotional intensity.

Speaker Familiarity

Though acoustic properties may influence children's understanding of vocal emotion, familiarity may also play an important role. Research suggests that infants respond significantly more to familiar rather than unfamiliar languages (Moon, Cooper, & Fifer, 1993; Kisilevsky et al., 2009) and rhymes (DeCasper, Lecanuet, Busnel, Granier-Deferre, & Maugeais, 1994). Of interest, however, is the significance of speaker familiarity, and in particular the mother's voice. For example, one study suggested that even fetuses are sensitive to their mother's voice compared to their father's or an unfamiliar female's voice (Kisilevsky et al., 2009). Further, infants as young as 7 months of age can better distinguish, based on listening time, their own mother's voice through background noise, relative to the same stimuli but with an unfamiliar female's voice (Barker & Newman, 2004). Despite evidence for the importance of the mother's voice, much of the existing literature for childhood emotion processing uses unfamiliar voices or faces as stimuli. Some research suggests that familiarity with a person influences children's understanding of facial emotional expressions (Herba et al., 2008; Montague & Walker-Andrews, 2002). To our knowledge, only one study demonstrated that children aged 7- to 12-years-old more accurately labeled their mothers' vocal emotion than that of an unfamiliar female, and that this effect differed depending on the emotion being conveyed (Shackman & Pollak, 2005). Further research is needed to consider how speaker familiarity and the emotion being conveyed may affect children's abilities to interpret prosodic emotion cues. The present study aimed to address this question by including mothers and unfamiliar mothers as the vocal stimuli.

Mothers, compared to non-mothers, have different acoustical profiles in speech (Kempe, Schaeffler, & Thoresen, 2010) and alterations to their vocal chords as a result of pregnancy (Cassiraga, Castellano, Abasolo, Abin, & Izbizky, 2012). The present study used mothers for both familiar and unfamiliar groups to further establish if differences based on speaker familiarity are specific to a child's own mother rather than just any mother compared to a non-mother.

The Present Study

The present study addresses the unique and joint roles of speaker acoustic properties, emotion, and speaker familiarity in 7- and 8-year-olds' ratings of the intensity of different affective prosodies while listening to natural speech samples. Specifically, we investigated the acoustic properties associated with three basic emotions—anger, sadness, and happiness—using mean F0, F0 standard deviation, and speech rate, respectively, and speaker familiarity—mother versus an unfamiliar mother—in relation to children's ratings of emotional intensity. Moving beyond studies of children's accuracy in labeling personally irrelevant experimental stimuli, this study addresses the determinants of children's emotion intensity ratings when the stimuli vary in personal relevance. We created stimuli that were more ecologically valid than standard experimental stimuli in the sense that our stimuli were designed to better mimic how children hear and use acoustic information in their daily lives by using natural speech. Additionally, we varied speaker familiarity, inviting children's own mothers to produce speech stimuli and comparing those stimuli to those produced by other mothers in the study who were unfamiliar to the child. Further, we focused on children's intensity ratings of affective prosody rather than accuracy in order to capture the more dynamic nature of the human voice.

This study aimed to 1) determine if F0, F0 standard deviation, and speech rate influence children's ratings of emotional intensity for angry, happy, and sad prosodies, 2) examine if children's ratings of emotional intensity differ based on speaker or emotion, and 3) determine the relative influence of speaker familiarity on children's emotional intensity ratings compared to the influence of the acoustic properties. We hypothesized that 1) variations of F0, F0 standard deviation, and speech rate would predict emotional intensity ratings. Though the literature suggests specific acoustical profiles using these properties for recognition of each emotion in research with adults (Pell et al., 2009), we did not have any directional hypotheses because of the limited literature with child participants. Based on the literature examining the importance of familiarity on children's emotion recognition (Shackman & Pollak, 2005), we also hypothesized that 2) children would rate their own mothers as more intensely emotional than the unfamiliar mothers, and in particular for anger. Finally, we hypothesized that 3) speaker familiarity would account for significant variance in children's emotional intensity ratings for all emotions even after accounting for the influence of the acoustic parameters.

Chapter 2. METHOD

Participants

Participants included 54 mothers (*M*age = 39.09 years, SD = 4.78 years) and their typically developing children aged between 7 years 0 months and 8 years 11 months recruited from a small mid-Atlantic city for inclusion in the multi-visit study. All mothers provided vocal stimuli for inclusion in the study. Six mothers were not included as unfamiliar stimuli: 3 mothers had strong accents that may have prosodic (Anderson-Hsieh & Koehler, 1988) and speech rate (van Maastricht, Krahmer, & Swerts, 2016) differences compared to native speakers or negatively impact children's speech comprehension (Munro & Derwing, 1995; Bent & Atagi, 2017), and 3 mothers were recorded near the end of data collection when no more unfamiliar mother voices could be used. As a result, 48 mothers served as the voice of an unfamiliar mother for another child in the study. A different set of 3 mothers were not included as familiar voices because they chose not to participate after the recording visit, resulting in a total of 51 mothers serving as the voice of a familiar mother. 45 mothers appeared in both groups. Child participants included a final sample of 51 children (49% female; M_{age} = 8.03 years, SD_{age} = 0.51 years) who completed all study visits.

This study was reviewed and approved by the university Institutional Review Board. All participants provided informed consent, and all children provided assent before participation.

Procedure

The data for this study came from a larger study involving four data collection points. At the first visit, mothers came to the lab and recorded the vocal stimuli that children heard while in the scanner during the second visit. While mothers were recording, children completed two subtests of the Diagnostic Analysis of Nonverbal Accuracy (DANVA-II; Nowicki & Duke, 1994), a computerized assessment of nonverbal emotion (faces and voices) and the standardized Clinical Evaluation of Language Fundamentals (CELF-5; Wiig, Secord, & Semel, 2013), to insure they had no receptive language delay. Finally, the children also had an orientation to fMRI scanning, including practice remaining still. After these activities were completed, mother and child were escorted to the imaging center for a mock fMRI scan to further prepare the children for the second visit. During the second visit, children were asked to complete a second mock scan and then the actual scan during which they heard the mothers' recordings for the first time. Between the second and fourth visits, children wore an iPod equipped with an electronically activated recording program that randomly sampled sounds in the children's home environments over the course of one day.

The fourth and final data collection occurred during a home visit. During that visit, children listened to the speech stimuli that they had heard in the scanner in order to evaluate their perceptions of the speech, including the intensity of the emotion conveyed. All speech stimuli were presented on a laptop through headphones. Children rated each individual recording in multiple ways, including the information used for this study. The children were asked "How (angry/happy/sad/scared) does this sound?" and then rated the emotional intensity of each speech sample for the presence of all four emotions on a scale from 1 (not) to 4 (a lot), with a visual aid to help them calibrate intensity (see Figure 1).

Figure 1. *Visual Aid to Assist Children with Emotional Intensity Ratings*.

In addition, children rated the voices for familiarity as well as for how the stimuli made the children feel. The present study focuses only on children's ratings of emotional intensity for the target emotion of the voices and the acoustic properties of the speech stimuli, described next.

Stimuli Creation

To create the speech stimuli, each mother completed a training and recording session at the first visit. The research assistant knowledgeable about affective prosody began with an introduction to how the voice communicates emotion, using descriptive terms taken from the emotion literature (Table 1) and clips from rehearsals for the film “Inside Out” (Pixar Animation Studios, 2015), in which actors attempted to convey emotions in their voices. The mother then was given an opportunity to discuss the task, to recall times when they felt the target emotions, and to practice generating emotion in their voices.

Table 1. *Descriptors of the Target Affects Provided to Mothers before Recording*.

Target Affect			
Non-Emotional	Anger	Sadness	Happiness
removed	sharp	subdued	melodic
complacent	biting	low-energy	sing-song
apathetic	harsh	exasperated	chirpy
robotic	abrupt	lethargic	sprightly
flat	frontal	depressed	chipper
factual	raised	exhausted	liltingly
news-report	demanding	hopeless	quick
	coarse	breathy	hopeful

Mothers were next shown the eight semantically non-emotional scripts to review and rehearse before recording began. The vocal coach corrected any pronunciation or semantic errors prior to the start of recording. Each script was designed as one side of a telephone conversation with another adult (see Appendix for all scripts). Mothers recorded scripts in a randomized order, but with a standardized order of four prosodies: non-emotional, angry, sad, and happy. As a result, each mother produced 32 stimuli of 8 scripts in each of 4 prosodies.

Recording took place in a sound-isolated room (Figure 2). A graduate student in the Acoustics department served as the sound engineer in the adjacent room. The recording studio had rubber spacers below the floor and between the walls and an approximately 6-inch thick door to further minimize sound transfer. Though there was an air vent that produced some noise, background noise was minimized by orienting the microphone in the farthest corner and turned away from the air vent. Recording equipment included a condenser microphone and shock mount to prevent vibrations from the microphone stand transferring during recording. The sound engineer also provided feedback to the mother to ensure the highest quality recordings.

Figure 2. *Picture of the Recording Studio*.

Each mother completed one practice and two useable recordings with feedback between takes for each script spoken in each of the four prosodies. The sound engineer selected the best sample for each script in each prosody.

In preparation for the scan, the sound engineer created the individualized stimuli for each child using the Digital Audio Workstation (DAW) in the Adobe Audition software (Adobe Systems, 2015) for recording and MATLAB (The Mathworks Inc., 2005) for processing. The stimuli were standardized to all be exactly 10 seconds in length by inserting or removing silences between phrases. The stimuli were also standardized for loudness (amplitude) so that each could be heard at the same volume in the noisy MRI environment. This standardization entailed applying the A-weighted root mean square (RMS) for amplitude for each recording to match the stimulus with the lowest RMS value. A-weighting of the amplitudes mimics the perceptual range of amplitudes of the human ear to make sure that “unheard” sounds do not confound the loudness standardization. This standardization also eliminates the possibility that differences in children’s emotional intensity ratings would be due to the loudness of a particular stimulus. The specific recordings used are available to authorized researchers via Databrary (Cole et al., 2017).

At both the scan session during which the children heard the stimuli for the first time and the home visit during which children rated the stimuli, children heard all 32 stimuli recorded by their own mother and all 32 stimuli recorded by a randomly selected unfamiliar mother, resulting in a total of 64 stimuli for each child. Using an unfamiliar mother as each child’s control voice ensured that the familiar and unfamiliar groups had the same variability in production of affective prosody. This method ensured that the individual variations in emotional expressivity were equivalent at the group level.

Affective Prosody Variables: Children’s Ratings and Acoustic Properties

Intensity Ratings. Children heard and provided emotional intensity ratings for 64 recordings: 8 non-emotional, 8 angry, 8 happy, and 8 sad recordings each for 2 speakers. The present study focuses on the angry, happy, and sad prosodies. As noted, children rated each recording on four emotions, but the present study only analyzed children’s intensity ratings of the target emotions. Emotional intensity ratings were averaged for each emotion by speaker to create intensity rating variables, resulting in 6 scores for each child—3 prosodies (angry, happy, and sad) and 2 speakers (own and unfamiliar mother).

Acoustic Parameters. Acoustic parameters were curated for each recording using the speech processing software, PRAAT (Boersma & Weenink, 2001; https://praat.org), to acquire numeric representations for each target acoustic parameter. The sound engineer (Acoustics graduate student) created individual code sequences (i.e., scripts) in PRAAT. The present study analyzed the acoustic parameters for the angry, happy, and sad recordings. The target acoustic parameters were:

F0 Mean. Fundamental Frequency (F0) mean was calculated for each stimulus using an autocorrelation algorithm (Boersma, 1993). In essence, F0 mean represents the average pitch or mid-point of the frequencies present in a given vocal track. The F0 means for all recordings within a specific prosody and by a specific speaker were averaged together, resulting in six separate F0 mean values for each child (3 emotions x 2 speakers).

F0 SD. F0 standard deviation was calculated by taking the square root of the variance of the pitches within a recording. This is a measure of how spread out or variable the pitches are in a given recording. The F0 standard deviations for all recordings within a specific prosody and by a specific speaker were averaged together, resulting in six separate F0 standard deviation values for each child (3 emotions x 2 speakers).

Syllabic Rate. Syllabic rate was calculated by dividing the number of syllables within a phrase by the spoken duration of that phrase to determine number of syllables per second. This is a measure of how quickly or slowly the speaker says the words. These were established based on the number of syllables in the text of each standardized script and the individual phrase durations created by the participants. Syllabic rates for all recordings within a specific prosody and by a specific speaker were averaged together, resulting in six separate syllabic rate values for each child (3 emotions x 2 speakers).

Data Analysis

All analyses were completed using the IBM Statistical Package for the Social Sciences (SPSS) for Windows, version 24. Descriptive statistics for the three acoustic parameters and children's emotional intensity ratings are presented in Table 2. Before conducting analyses, the distributions of each variable were inspected for normality. The bivariate correlations are also presented in Table 3.

To establish if children accurately recognized the emotions conveyed in the recordings, we used a series of one-way Analyses of Variance (ANOVAs), one for each target emotion. Specifically, we compared the average target emotional intensity rating to the average emotional intensity ratings of the other three emotions on which children rated all recordings. For example, an ANOVA for angry recordings tested the average angry intensity rating to the average happy, average sad, and average scared intensity ratings of those same recordings. For these angry recordings, children would be considered accurate if they rated the average angry intensity significantly higher than the intensity of happy, sad, or scared for the angry recordings.

To test the hypothesis that acoustic properties influence children's emotional intensity ratings, we used three linear regressions, one for each of the target emotions (angry, happy, and sad). Specifically, we tested whether F0, F0 standard deviation, and syllabic rate as a set contribute significantly to variance in children's emotional intensity ratings for each emotion. The predictors were F0, F0 standard deviation, and syllabic rate, all included in one step, and the outcome was the intensity rating for that specific emotion.

To test the hypothesis that children rate their mothers' speech stimuli as more intensely emotional than the stimuli produced by unfamiliar mothers, we conducted a 2 (speaker familiarity) X 3 (emotion) Multivariate Analysis of Variance (MANOVA). Children's intensity ratings were the dependent variables. Planned comparisons were used to test whether there were significant differences in speaker familiarity as a function of emotion.

Finally, to test the hypothesis that speaker familiarity accounts for emotional intensity ratings, over and above acoustic properties, we conducted three hierarchical regressions. The acoustic variables—mean F0, F0 standard deviation, and syllabic rate—were entered together in the first step and speaker familiarity was entered as a contrast in the second step. Children's intensity ratings for each emotion were the dependent variables.

Chapter 3. RESULTS

Descriptive Statistics

Means and standard deviations for the emotional intensity ratings, F0 Mean, F0 SD, and Speech Rate for the angry, happy, and sad stimuli are included in Table 2. Emotional intensity ratings were based on a scale from 1 (e.g., “Not Angry”) to 4 (e.g., “A lot Angry”). Children’s average intensity ratings for each emotion of approximately 3 (e.g., “Some Angry”) indicated that children perceived the emotions as moderately intense. Correlation analyses indicated that emotional intensity ratings were not significantly associated with any of the acoustic variables for any emotion category (Table 3). The F0 Mean was positively correlated with F0 SD for all three emotion categories, all *p*s < .001, and negatively correlated with Speech Rate only for Sad speech stimuli, $r = -.34$, $p = .016$.

Table 2.

Descriptive Statistics for the Variables of Interest for Angry, Happy, and Sad Prosodies

	Angry		Happy		Sad	
	M	SD	*M*	SD	*M*	SD
Intensity Rating	3.16	0.46	3.10	0.61	2.96	0.63
F0 Mean	229.60	24.03	318.09	22.83	183.66	12.73
F0 SD	49.54	7.30	93.84	14.15	22.73	3.00
Speech Rate	3.83	0.28	4.20	0.42	3.40	0.29

Note. F0 = Fundamental Frequency. Intensity ratings were on a scale from 1 – 4, with an average score of 3 representing “Some” of the target emotion.

Table 3.

Correlations for Variables of Interest for Angry, Happy, and Sad Prosodies for 51 Subjects

	1	2	3	4
Angry				
(1) Intensity Rating	-			
(2) F0 Mean	.008	-		
(3) F0 SD	-.102	.775***	-	
(4) Speech Rate	.035	-.057	-.138	-
Happy				
(1) Intensity Rating	-			
(2) F0 Mean	.124	-		
(3) F0 SD	.256	.609***	-	
(4) Speech Rate	.000	.123	-.101	-
Sad				
(1) Intensity Rating	-			
(2) F0 Mean	.154	-		
(3) F0 SD	-.097	.420**	-	
(4) Speech Rate	.045	-.336*	-.141	-

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. F0 = Fundamental Frequency.

Children's Emotion Identification Accuracy

A set of three one-way repeated measures ANOVAs, one for each target emotion, tested children's accuracy at recognizing the target emotion conveyed in the recordings. Results revealed that children rated angry recordings as being more intensely angry than those recordings were happy, sad, or scared (all *p*s < .001). Children rated happy recordings as being more intensely happy than those recordings were angry, sad, or scared (all *p*s < .001). Finally, children rated sad recordings as being more intensely sad than those recordings were angry, happy, or scared (all *p*s < .001).

Acoustic Properties Predicting Emotional Intensity

The first set of three regression analyses, one for each emotion, tested the hypothesis that varying levels of the acoustic properties (F0, F0 SD, and speech rate) would predict children's emotional intensity ratings of the speech stimuli, regardless of speaker familiarity. All variables were included in one step in the model. Contrary to the hypothesis, the acoustic properties did not predict the emotional intensity ratings for any of the target emotions (angry, happy, or sad), all *t*s < 1.7, all *p*s > .100. Full results for each of the regression models are outlined in Table 4.

Table 4.

Results of Regression Analyses of Acoustic Properties Predicting Children's Emotional Intensity Ratings Collapsed Across Speaker.

	B	*t*	*p*	η_p^2	*F*	*df*	*p*	R^2
Angry								
Overall Model				.03	0.48	3, 47	.701	0.03
F0 Mean	.004	0.95	.347	.02				
F0 SD	-.017	-1.17	.249	.03				
Speech Rate	.017	0.07	.945	<.01				
Happy								
Overall Model				.07	1.15	3, 47	.340	0.07
F0 Mean	-.002	-0.34	.739	<.01				
F0 SD	.013	1.63	.110	.05				
Speech Rate	.055	0.25	.801	<.01				
Sad								
Overall Model				.07	1.10	3, 47	.358	0.07
F0 Mean	.013	1.67	.102	.06				
F0 SD	-.041	-1.26	.213	.03				
Speech Rate	.233	0.73	.472	.01				

Note. F0 = Fundamental Frequency.

Differences in Emotional Intensity by Speaker or Emotion

A two-way repeated measures MANOVA model tested if children rated the emotional intensity of the speech stimuli differently based on the speaker, the emotion, or their interaction (Table 5). For emotion, there were no significant findings but, given the novelty of this type of research, it may be noteworthy that there was a trend towards a main effect of emotion, $F(2, 49) = 2.97$, $p = .061$, $\eta_p^2 = .11$. Though there was no directional hypothesis for differences in intensity ratings by emotion, post-hoc tests reviewed all pairwise comparisons using Bonferroni correction for multiple testing. The mean difference between angry ($M = 3.16$) and sad ($M = 2.96$) trended towards significance, $p = .058$, 95% CI [-0.01, 0.41]. Neither of the remaining pairwise comparisons were significant, all *p*s > .310.

Results also suggested a main effect of speaker familiarity. Consistent with our hypothesis, children on average rated their mothers' voices ($M = 3.15$) as more intensely emotional than strangers' voices ($M = 2.99$), $F(1, 50) = 12.02$, $p = .001$, $\eta_p^2 = .16$, 95% CI [0.07, 0.25].

The main effect of speaker familiarity appeared to be moderated by emotion for children's intensity ratings, $F(2, 49) = 7.95$, $p = .001$, $\eta_p^2 = .25$, as seen in Figure 3. Simple effects tests revealed that children rated their own mothers ($M = 3.30$) as more intensely angry than the unfamiliar mothers ($M = 3.02$), $F(1, 50) = 14.25$, $p < .001$, $\eta_p^2 = .22$, 95% CI [0.13, 0.42]. Additionally, children rated their own mothers ($M = 3.21$) as more intensely happy than the unfamiliar mothers ($M = 2.99$), $F(1, 50) = 6.41$, $p = .015$, $\eta_p^2 = .11$, 95% CI [0.05, 0.40]. There was no significant effect of speaker familiarity for children's sad prosody ratings, $p = .712$.

Table 5.

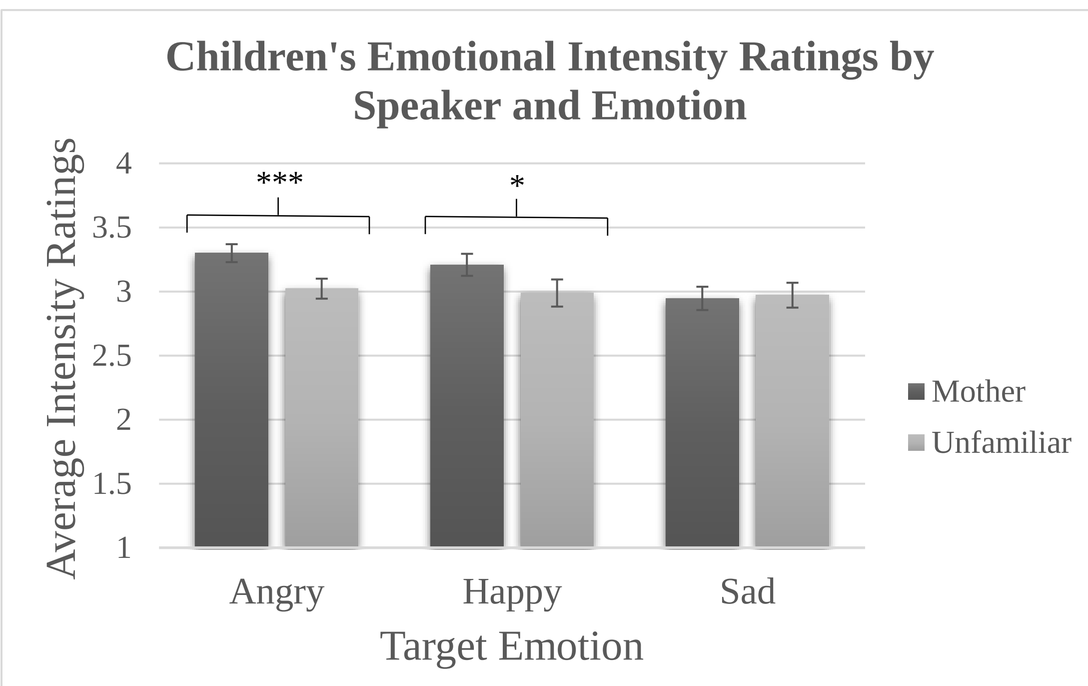
Results of MANOVA Testing for Difference in Children's Emotional Intensity Ratings by Emotion and by Speaker.

	F	*df*	*p*	η_p^2	Mean Diff.	*95% Confidence Interval* Lower Bound	Upper Bound
Emotion	2.97	2, 49	.061	.11			
Angry vs. Happy			1.00		0.06	-0.14	0.27
Angry vs. Sad			.058		0.20	-0.01	0.41
Happy vs. Sad			.319		0.14	-0.07	0.35
Speaker	12.02	1, 50	.001	.19	0.16	0.07	0.25
Emotion*Speaker	7.95	2, 49	.001	.25			
Within Angry	14.25	1, 50	<.001	.22	0.28	0.13	0.42
Within Happy	6.41	1, 50	.015	.11	0.22	0.05	0.40
Within Sad	0.14	1, 50	.712	<.01	-0.03	-0.16	0.11

Note. Bonferroni correction was used to determine significance of pairwise comparisons under the main effect of emotion. Mean Diff = mean difference between categories.

Figure 3.

Interaction of Children's Emotional Intensity Ratings by Speaker and Emotion.



Effect of Speaker Familiarity Over and Above Acoustics

Finally, three separate hierarchical regressions tested the hypothesis that speaker familiarity would predict children's emotional intensity ratings over and above the effects of the acoustic variables (Table 6). For angry stimuli, results indicated that speaker familiarity significantly predicted children's intensity ratings even after accounting for F0, F0 SD, and speech rate, $B = .14$, $t(97) = 2.69$, $p = .008$, $\eta_p^2 = .07$, 95% CI [0.04, 0.25]. Consistent with the hypothesis, speaker familiarity accounted for a significant additional amount of variance in emotional intensity ratings over and above that accounted for by the acoustic variables for angry stimuli, $\Delta R^2 = .07$, $F_\Delta(1, 97) = 7.25$, $p = .008$, $\eta_p^2 = .09$. For happy and sad stimuli, speaker familiarity did not significantly predict children's emotional intensity ratings, both *p*s > .110. Additionally, the inclusion of speaker familiarity in the model did not account for a significant amount of variance beyond that already accounted for by the acoustic variables, both *p*s > .150.

Table 6.

Results from Hierarchical Regression Analyses of Acoustic Properties and Speaker Predicting Children's Emotional Intensity Ratings.

	B	*t*	*p*	η_p^2	ΔR^2	*F*	*df*	*p*
Angry								
Step 1 Overall				.02	.02	0.79	3, 98	.500
F0 Mean	.003	1.24	.220	.02				
F0 SD	-.004	-0.49	.625	<.01				
Speech Rate	-.070	-0.50	.619	<.01				
Step 2 Overall				.09	.07	2.45	4, 97	.052
Speaker	.142	2.69	.008	.07				
Happy								
Step 1 Overall				.04	.04	1.41	3, 98	.244
F0 Mean	.001	0.37	.715	<.01				
F0 SD	.006	1.24	.220	.02				
Speech Rate	.039	0.33	.745	<.01				
Step 2 Overall				.07	.03	1.73	4, 97	.150
Speaker	.109	1.61	.110	.03				
Sad								
Step 1 Overall				.04	.04	1.23	3, 98	.301
F0 Mean	.008	1.84	.069	.03				
F0 SD	-.019	-1.19	.236	.01				
Speech Rate	.086	0.50	.615	<.01				
Step 2 Overall				.04	.00	0.92	4, 97	.453
Speaker	-.012	-0.18	.858	<.01				

Note. F0 = Fundamental Frequency.

Chapter 4. DISCUSSION

The goal of the present study was to examine the influences of both the acoustic properties of affective prosody (F0 mean, F0 SD, and speech rate) and speaker familiarity on children's ratings of emotional intensity for semantically neutral natural speech stimuli. Specifically, this study tested the hypothesis that the acoustic properties of angry, happy and sad natural speech would predict children's ratings of their intensity. Moreover, we tested whether factors other than the physical properties of affective prosody would influence children's ratings of perceived intensity. Specifically, we tested the hypothesis that children would rate their own mothers as more intensely emotional than unfamiliar mothers, and that this relation may depend on the emotion each mother expressed. Finally, we tested the hypothesis that speaker familiarity would account for variance in children's intensity ratings over and above the influence of the acoustic properties.

Contrary to our first hypothesis, F0 mean, F0 SD, and speech rate did not predict children's emotional intensity ratings for angry, happy, or sad stimuli. Prior research indicates that young children use the average pitch (mean F0) to identify joy and anger in the voice (Dmitrieva et al., 2008). Though these children accurately differentiated between emotions, the present findings yielded no evidence that acoustic parameters contribute to perceptions of emotion intensity. However, there are a few key differences between the aforementioned and present study to consider. First, the present study emphasized emotional intensity ratings in relation to acoustic properties rather than emotion labeling (Dmitrieva et al., 2008). Second, the present study used mothers with little to no acting experience, whereas the aforementioned study utilized one trained actor for all vocal stimuli (Dmitrieva et al., 2008). Actor training alters speakers' F0 range and mean F0 (Walzak, McCabe, Madill, & Sheard, 2008). However, there still appears to be no significant differences in acoustic properties between non-actor and actor portrayals of emotion (Jürgens, Grass, Drolet, & Fischer, 2015), therefore speaker portrayal of emotion is unlikely to explain why acoustic properties do not predict children's intensity ratings. In light of that information, our results may indicate that the acoustic cues that facilitate emotion labeling do not play a role in identifying emotional intensity. Instead, children may attend to other cues, like those psychological in nature, to interpret emotional intensity. More research is needed to understand what other factors may help children understand differences in intensity. This is further supported by our other results (as discussed later) that children rated the emotional intensity differently at the group level based on speaker familiarity, despite the fact that the two groups did not differ based on acoustic properties. However, it is difficult to know if children were able to accurately perceive differences in emotional intensity. The children in the present study accurately differentiated between emotion categories, but we do not know if children's perceptions of emotional intensity within an emotion category were consistent with actual variations in emotional intensity produced by the speakers. If they were not consistent, then it is possible that potential acoustic differences in intensity would not predict children's inaccurate perceptions of emotional intensity. With the goal of creating ecologically valid, natural speech samples, the present study did not control for variations in emotional intensity and instead allowed the speakers to speak as they believed they usually would. Further research is needed to control for variations in relative intensity to establish if children can understand differences in emotional intensity within the same emotion category, though this would come at the cost of more natural speech stimuli.

The results supported our second hypothesis. Children rated their own mothers as more intensely emotional compared to unfamiliar mothers, an overall effect that was moderated by the emotion being heard, specifically for angry and happy, but not sad, prosodies. These results are in-line with the literature suggesting that infants and children attend more to positive and negative emotions expressed by a familiar person, their mothers in particular (Kahana-Kalman & Walker-Andrews, 2001; Montague & Walker-Andrews, 2002; Shackman & Pollak, 2005; Todd, Evans, Morris, Lewis, & Taylor, 2011; Arsalidou, Barbeau, Bayless, & Taylor, 2010). Children in the present study may have perceived their mothers as more intensely emotional than unfamiliar mothers because they were attending more to their mothers' voices in particular. Given that 7- and 8-year-olds have spent a significant amount of time with their mothers through their young lives, their mothers' emotional states will have more direct and immediate consequences. Children may be intrinsically motivated to have a higher sensitivity for their mothers' emotional states, as opposed to those of unfamiliar women, in order to better anticipate potential outcomes in their daily lives, particularly for when their mothers are angry or happy. Alternatively, it is possible that children rated their mothers as more intensely angry and more intensely happy than unfamiliar mothers because children are better able to discern differences in their mothers' states. For example, throughout their daily lives, children will experience the entire range of their mothers' emotions, from elation to passive amusement, and from disappointment to hot anger. Children listening to an unfamiliar woman, however, will not know the possible variations in her emotions and therefore may not know the difference between her intense anger and mild anger. This effect goes beyond the result of changes that may occur as a result of pregnancy (Cassiraga et al., 2012) or motherhood (Kempe et al., 2010). The present study used approximately the same group of mothers as both the familiar and unfamiliar voices, eliminating this possibility. As a result, differences in these children's ratings of emotional intensity for anger and happiness for *their own* mother compared to another mother were entirely perceptual. This design, in conjunction with the literature, highlights the significance of the mother in children's daily lives and how these daily experiences influence the development of emotional awareness.

Finally, in partial support of our third hypothesis, the effect of speaker familiarity emerged over and above any influences of the acoustic properties only for the angry speech stimuli. In other words, speaker familiarity played a more important role than acoustic properties for children's emotional intensity ratings of specifically anger. This is consistent with other research using event-related potentials (ERPs) to assess the neural responses to emotion in the face as a function of familiarity among typically developing 4- to 6-year-olds that reported that children exhibited a significantly longer processing latency following mothers' angry faces, but not their happy faces or an unfamiliar woman's face, suggesting that children spent more time processing their mothers' angry faces (Todd, Lewis, Meusel, & Zelazo, 2008). Additionally, vocal emotion research has suggested that both typically developing and physically abused children focus more on angry vocal expressions in their mothers' voices even more so than the angry expressions of an unfamiliar female (Shackman & Pollak, 2005). Overall, these data suggest that anger may be unique in children's perceptions of emotion such that the person who is angry may be useful information than the quality of the anger itself. They underscore the fact that emotional understanding of anger may be less about the physical properties of voices and more about *who* is angry. Further, it is unclear if this effect may be situational. For example, children in the present study listened to contextually non-emotional adult-directed speech such that their mothers or an unfamiliar mother were speaking over the phone to an unknown individual. For many children, in a novel research situation, an unfamiliar woman's angry voice *not* directed at them may not be meaningful. Their own mothers' angry voice, even directed at another individual, is salient in a novel context, particularly when both voices are being heard in quick succession. There may be situations, however, where an unfamiliar woman's anger directed at another individual may still impact children's actions, for example if that child is being taught to do something that makes that unfamiliar woman angry (Repacholi & Meltzoff, 2007) and their own mother is not speaking. Additional research is necessary to clarify under what circumstances speaker familiarity more strongly influences children's interpretation of emotional intensity above other properties.

Finally, exploratory post-hoc analyses of the effect of emotion on intensity ratings suggested a potential difference in children's intensity ratings for angry and sad recordings. This means that children perceived angry recordings as being more intensely angry than they perceived the sad recordings as intensely sad. One interpretation of these results is that the differences in intensity ratings for angry and sad prosodies reflect the actual acoustical differences in the two emotions (Banse & Scherer, 1996; Jallais & Gilet, 2010; Yildirim et al., 2004). More specifically, vocal anger requires higher levels of arousal and energy, which are both acoustic indicators of intensity, while sadness has the lowest levels (Banse & Scherer, 1996; Jallais & Gilet, 2010; Yildirim et al., 2004). The results of this study then suggest that children in fact may actually be accurately perceiving the emotional intensities of anger and sadness, in direct contrast to the prior interpretation. Another possible interpretation of this finding could be that the children may have had more difficulty understanding and identifying sadness than anger. This would be consistent with literature suggesting that at least for facial emotion, children's abilities to identify sadness increase with age, but they are consistently able to identify anger (Herba et al., 2008). For the present study, the participants' ratings of sadness would be of lower intensity ("none" or "a little") if they were unable to differentiate sadness from other emotions.

Angry recordings may have been more easily identifiable and therefore children may have had an easier time identifying intense anger. However, given that anger and sadness are acoustically different, it is more likely that children were accurately perceiving these differences. Further, that their intensity ratings are consistent with acoustical differences between anger and sadness also suggests that children may be attending to those other acoustic properties, arousal and energy of speech, to determine emotional intensity rather than mean F0, F0 standard deviation, and speech rate. Further research is needed to directly investigate how these other acoustical factors may influence children's perceptions of emotional intensity, as well as what downstream effects those perceptions may have on their behavior.

Limitations and Future Directions

These results should be considered in light of several limitations, as previously noted. First, the relation between emotion identification and the specific acoustic properties included in this study are supported by a wide literature. However, that literature is limited to emotion labeling and a few consistently identified acoustic properties. Future research should examine other acoustic properties like mean energy to establish if other properties of the voice may contribute to perception of emotional intensity specifically, rather than emotion labeling. As previously noted, children perceived anger and sadness as different levels of intensity, and both emotions tend to differ on mean energy (Banse & Scherer, 1996; Jallais & Gilet, 2010; Yildirim et al., 2004). Second, one strength of the design of this study involved children rating each recording on the intensity of four separate emotions, which acknowledges the dynamic and nuanced nature of emotion expression. However, we only focused on the rating for the target emotion to allow for sufficient power in the analyses. Future research should consider that one speech sample can convey multiple emotions, and test if the perception of multiple emotions alters the relationship between the acoustic properties of those samples and emotional intensity ratings. Third, this study examined a very limited age range that exhibits an unclear ability to accurately identify emotions in the voice throughout the literature. Future research could follow children longitudinally to help elucidate if and when acoustic properties influence their abilities to understand emotional intensity, as well as what factors affect how this ability develops as children age. Finally, the present study consolidated and averaged ratings and acoustic information from multiple stimuli in order to accommodate the analytic approach. Future studies could employ a more sophisticated statistical approach like Multilevel Modeling that would account for the repeated and nested measurements while still capturing the variability in acoustic properties for each separate recording and how that may influence differences in children's emotional intensity ratings.

Conclusions

Overall, this study examined children's perceptions of emotional intensity of angry, happy, and sad speech in relation to the acoustic properties of the mean fundamental frequency, fundamental frequency standard deviation, and speech rate, and speaker familiarity. Results emphasized the importance of speaker familiarity in children's interpretations of intensity in emotional speech, particularly for angry prosody. These results also suggested that children may not be attending to the acoustic properties in speech to influence their perceptions of emotional intensity. Instead, differences in perceived intensity may be driven by more psychological factors related to speaker familiarity.

APPENDIX

Natural speech scripts that all mothers recorded and children rated for emotional intensity.

Script Topic	Version	Text
Checkbook	A	"Where is the checkbook? It's gone, I can't find it. I don't have it. Do you have it?"
	B	"Do you have the checkbook? You had it last. It's just not here. I'll look for it."
Dinner	A	"Oh, hi, it's you. When will you be home? Dinner won't be ready then. Okay, I'll fix dinner."
	B	"I'm fixing dinner. It will take an hour. I have a lot to do. I'll see you later."
Help	A	"Hi, I hoped you'd call. You're running late? I will need some help. Can you change your plans?"
	B	"I could use your help. There's so much to do. Can you change your plans? See you when you get here."
Talk	A	"Oh, you're tired? Sorry to hear that. We should talk. About lots of things."
	B	"Can you talk now? About lots of things. Money, the weekend. Okay, we won't talk now."

Note. Mothers recorded each script in the angry, happy, sad, and non-emotional prosodies.

References

Adobe Systems. (2015). Adobe Audition [Computer Program]. San Jose: Adobe. Retrieved from https://www.adobe.com/Audition

Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology*, *12*(2), 169-177. doi: 10.1016/S0959-4388(02)00301-X

Aguert, M., Laval, V., Lacroix, A., Gil, S., & Le Bigot, L. (2013). Inferring emotions from speech prosody: not so easy at age five. *PLoS ONE*, *8*(12), e83657. doi: 10.1371/journal.pone.0083657

Anderson-Hsieh, J., & Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning*, *38*(4), 561-613. doi: 10.1111/j.1467-1770.1988.tb00167.x

Arsalidou, M., Barbeau, E. J., Bayless, S. J., & Taylor, M. J. (2010). Brain responses differ to faces of mothers and fathers. *Brain and Cognition*, *74*(1), 47-51. doi: 10.1016/j.bandc.2010.06.003

Bachorowski, J., & Owren, M. J. (2003). Sounds of emotion. *Annals of the New York Academy of Sciences*, *1000*(1), 244-265. doi: 10.1196/annals.1280.012

Baltaxe, C. A. M. (1991). Vocal communication of affect and its perception in three- to four-year-old children. *Perceptual and Motor Skills*, *72*(3), 1187-1202. doi: 10.2466/pms.1991.72.3c.1187

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614-636. doi: 10.1037/0022-3514.70.3.614

Bänziger, T., & Scherer, K. R. (2005). The role of intonation in emotional expressions. *Speech Communication*, *46*(3-4), 252-267. doi: 10.1016/jspecom.2005.02.016

Barker, B. A., & Newman, R. S. (2004). Listen to your mother! The role of talker familiarity in infant streaming. *Cognition*, *94*(2), B45-B53. Doi: 10.1016/j.cognition.2004.06.001

Bent, T., & Atagi, E. (2016). Perception of nonnative-accented sentences by 5- to 8-year-olds and adults: the role of phonological processing skills. *Language and Speech*, *60*(1), 110-122. doi: 10.1177/0023830916645374

Berman, J. M. J., Chambers, C. G., & Graham, S. A. (2010). Preschoolers' appreciation of speaker vocal affect as a cue to referential intent. *Journal of Experimental Child Psychology*, *107*(2), 87-99. doi: 10.1016/j.jecp.2010.04.012

Berman, J. M. J., Graham, S. A., & Chambers, C. G. (2013). Contextual influences on children's use of vocal affect cues during referential interpretation. *The Quarterly Journal of Experimental Psychology*, *66*(4), 705-726. doi: 10.1080/17470218.2012.713367

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences*, *17*(1193), 97-110.

Boersma, P., & Weenink, D. (2001). PRAAT (version 4.0) [Computer software]. Retrieved from http://www.praat.org/

Breitenstein, C., Van Lancker, D., & Daum, I. (2001). The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample. *Cognition and Emotion*, *15*(1), 57-79. doi: 10.1080/02699930126095

Cassiraga, V. L., Castellano, A. V., Abasolo, J., Abin, E. N., & Izbizky, G. H. (2012). Pregnancy and voice: changes during the third trimester. *Journal of Voice*, *26*(5), 584-586. doi: 10.1016/j.jvoice.2011.10.004

Cole, P.M., Gilmore, R.O., Perez-Edgar, K. & Scherf, K.S. (2017). Children's Neural Processing of the Emotional Environment (PEEP-II). Databrary. Retrieved August 29, 2018 from https://nyu.databrary.org/volume/339.

DeCasper, A. J., Lecanuet, J., Busnel, M., Granier-Deferre, C., & Maugeais, R. (1994). Fetal reactions to recurrent maternal speech. *Infant Behavior and Development*, *17*(2), 159-164.

Dmitrieva, E. S., Gel'man, V. Y., Zaitseva, K. A., & Orlov, A. M. (2008). Dependence of the perception of emotional information of speech on the acoustic parameters of the stimulus in children of various ages. *Human Physiology*, *34*(4), 527-531. doi: 10.1134/S036211970804018X

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology*, *27*(2), 209-221. doi: 10.1037/0012-1649.27.2.209

Fernald, A. (1993). Approval and disapproval: infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development, 64*(3), 657-674. doi: 10.1111/j.1467-8624.1993.tb02934.x

Friend, M. (2000). Developmental changes in sensitivity to vocal paralanguage. *Developmental Science, 3*(2), 148-162. doi: 10.1111/1467-7687.00108

Friend, M. (2003). What should I do? Behavior regulation by language and paralanguage in early childhood. *Journal of Cognition and Development*, *4*(2), 161-183. doi: 10.1207/S15327647JCD0402_02

Friend, M., & Bryant, J. B. (2000). A developmental lexical bias in the interpretation of discrepant messages. *Merrill Palmer Q (Wayne State University Press), 46*(2), 342-369.

Halberstadt, A. G., Denham, S. A., & Dunsmore, J. C. (2001). Affective social competence. *Social Development*, *10*(1), 79-119. doi: 10.1111/1467-9507.00150

Herba, C. M., Benson, P., Landau, S., Russell, T., Goodwin, C., Lemche, E., Santosh, P., & Phillips, M. (2008). Impact of familiarity upon children's developing facial expression recognition. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *49*(2), 201-210. doi: 10.1111/j.1469-7610.2007.01835.x

Herba, C. M., Landau, S., Russell, T., Ecker, C., & Phillips, M. L. (2006). The development of emotion-processing in children: effects of age, emotion, and intensity. *Journal of Child Psychology and Psychiatry*, *47*(11), 1098-1106. doi: 10.1111/j.1469-7610.2006.01652.x

Jallais, C., & Gilet, A. (2010). Inducing changes in arousal and valence: comparison of two mood induction procedures. *Behavior Research Methods*, *42*(1), 318-325. doi: 10.3758/BRM.42.1.318

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*(3), 195-214. doi: 10.1007/s10919-015-0209-5

Juslin, P. N., & Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion*, *1*(4), 381-412.

Kahana-Kalman, R., & Walker-Andrews, A. S. (2001). The role of person familiarity in young infants' perception of emotional expressions. *Child Development*, *72*(2), 352-369. doi: 10.1111/1467-8624.00283

Kempe, V., Schaeffler, S., & Thoresen, J. C. (2010). Prosodic disambiguation in child-directed speech. *Journal of Memory and Language*, *62*(2), 204-225. doi: 10.1016/j.jml.2009.11.006

Kisilevsky, B. S., Hains, S. M. J., Brown, C. A., Lee, C. T., Cowperthwaite, B., Stutzman, S. S., Swansburg, M. L., Lee, K., Xie, X., Huang, H., Ye, H. H., Zhang, K., & Wang, Z. (2009). Fetal sensitivity to properties of maternal speech and language. *Infant Behavior and Development*, *32*(1), 59-71. doi: 10.1016/j.infbeh.2008.10.002

The Mathworks, Inc. (2005). MATLAB (Version 7.0) [Computer software]. Natick, Massachusetts: Mathworks.

McElwain, N. L., Halberstadt, A. G., & Volling, B. L. (2007). Mother- and father-reported reactions to children's negative emotions: Relations to young children's emotional understanding and friendship quality. *Child Development*, *78*(5), 1407-1425. doi: 10.1111/j.1467-8624.2007.01074.x

Montague, D. P. F., & Walker-Andrews, A. S. (2002). Mothers, fathers, and infants: the role of person familiarity and parental involvement in infants' perception of emotion expression. *Child Development*, *73*(5), 1339-1352. doi: 10.1111/1467-8624.00475

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, *16*(4), 495-500.

Morton, J. B., & Trehub, S. E. (2001). Children's understanding of emotion in speech. *Child Development*, *72*(3), 834-843. doi: 10.1111/1467-8624.00318

Morton, J. B., Trehub, S. E., & Zelazo, P. D. (2003). Sources of inflexibility in 6-year-olds' understanding of emotion in speech. *Child Development, 74*(6), 1857-1868. doi: 10.1046/j.1467-8624.2003.00642.x

Mumme, D. L., Fernald, A., & Herrera, C. (1996). Infants' responses to facial and vocal emotional signals in a social referencing paradigm. *Child Development, 67*(6), 3219-3237. doi: 10.1111/j.1467-8624.1996.tb01910.x

Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 38*(3), 289-306. doi: 10.1177/002383099503800305

Nelson, N. L., & Russell, J. A. (2011). Preschoolers' use of dynamic facial, bodily, and vocal cues to emotion. *Journal of Experimental Child Psychology*, *110*(1), 52-61. doi: 10.1016/j.jecp.2011.03.014

Nowicki, S., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: the Diagnostic Analysis of Nonverbal Accuracy Scale. *Journal of Nonverbal Behavior*, *18*(1), 9-35. doi: 10.1007/BF02169077

Nygaard, L. C., & Queen, J. S. (2008). Communicating emotion: linking affective prosody and word meaning. *Journal of Experimental Psychology: Human Perception and Performance, 34*(4), 1017-1030. doi: 10.1037/0096-1523.34.4.1017

Paulmann, S., Titone, D., & Pell, M. D. (2012). How emotional prosody guides your way: evidence from eye movements. *Speech Communication, 54*(1), 92-107. doi: 10.1016/j.specom.2011.07.004

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*(4), 417-435. doi: 10.1016/j.wocn.2009.07.005

Pixar Animation Studios. (2015). *Inside Out Behind the Scenes Footage – Amy Poehler, Bill Hader, Mindy Kaling, Lewis Black* [Online Video]. Available from https://www.youtube.com/watch?v=6YiqKqgd_jU&t=1s

Pollak, S. D., & Sinha, P. (2002). Effects of early experience on children’s recognition of facial displays of emotion. *Developmental Psychology*, *38*(5), 784-791. doi: 10.1037/0012-1649.38.5.784

Quam, C., & Swingley, D. (2012). Development in children’s interpretation of pitch cues to emotions. *Child Development*, *83*(1), 236-250. doi: 10.1111/j.1467-8624.2011.01700.x

Repacholi, B. M., & Meltzoff, A. N. (2007). Emotional eavesdropping: infants selectively respond to indirect emotional signals. *Child Development*, *78*(2), 503-521. doi: 10.1111/j.1467-8624.2007.01012.x

Rothman, A. D., & Nowicki, S. Jr. (2004). A measure of the ability to identify emotion in children’s tone of voice. *Journal of Nonverbal Behavior*, *28*(2), 67-92. doi: 10.1023/B:JONB.0000023653.13943.31

Saarni, C. (1999). *The development of emotional competence.* New York, NY: The Guilford Press.

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

Schachner, A., & Hannon, E. E. (2011). Infant-directed speech drives social preferences in 5-month-old infants. *Developmental Psychology*, *47*(1), 19-25. doi: 10.1037/a0020740

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*(3), 269-285. doi: 10.1007/BF01074443

Shackman, J. E., & Pollak, S. D. (2005). Experiential influences on multimodal perception of emotion. *Child Development*, *76*(5), 1116-1126. doi: 10.1111/j.1467-8624.2005.00901.x

Singh, L., Morgan, J. L., & Best, C. T. (2002). Infants' listening preferences: baby talk or happy talk? *Infancy*, *3*(3), 365-394. doi: 10.1207/S15327078IN0303_5

Sobin, C., & Alpert, M. (1999). Emotion in speech: the acoustic attributes of fear, anger, sadness, and joy. *Journal of Psycholinguistic Research*, *28*(4), 347-365. doi: 10.1023/A:1023237014909

Streeter, L. A., Macdonald, N. H., Apple, W., Krauss, R. M., & Galotti, K. M. (1983). Acoustic and perceptual indicators of emotional stress. *The Journal of the Acoustical Society of America*, *73*(4), 1354-1360. doi: 10.1121/1.389239

Todd, R. M., Evans, J. W., Morris, D., Lewis, M. D., & Taylor, M. J. (2011). The changing face of emotion: age-related patterns of amygdala activation to salient faces. *Social Cognitive and Affective Neuroscience*, *6*(1), 12-23. doi: 10.1093/scan/nsq007

Todd, R. M., Lewis, M. D., Meusel, L., & Zelazo, P. D. (2008). The time course of social-emotional processing in early childhood: ERP responses to facial affect and personal familiarity in a Go-Nogo task. *Neuropsychologia*, *46*(2), 595-613. doi: 10.1016/j.neuropsychologia.2007.10.011

Vaish, A., & Striano, T. (2004). Is visual reference necessary? Contributions of facial versus vocal cues in 12-month-olds' social referencing behavior. *Developmental Science*, *7*(3), 261-269. doi: 10.1111/j.1467.7687.2004.00344.x

Van Lancker, D., Cornelius, C., & Kreiman, J. (1989). Recognition of emotional-prosodic meanings in speech by autistic, schizophrenic, and normal children. *Developmental Neuropsychology*, *5*, 207-226. doi: 10.1080/87565648909540433

Van Maastricht, L., Krahmer, E., & Swerts, M. (2016). Native speaker perceptions of (non-)native prominence patterns: effects of deviance in pitch accent distributions on accentedness, comprehensibility, intelligibility, and nativeness. *Speech Communication*, *83*, 21-33. doi: 10.1016/j.specom.2016.07.008

Walker-Andrews, A. S., & Grolnick, W. (1983). Discrimination of vocal expressions by young infants. *Infant Behavior & Development*, *6*(4), 491-498. doi: 10.1016/S0163-6383(83)90331-4

Walzak, P., McCabe, P., Madill, C., & Sheard, C. (2008). Acoustic changes in student actors' voices after 12 months of training. *Journal of Voice, 22*(3), 300-313. doi: 10.1016/j.jvoice.2006.10.006

Waxer, M., & Morton, J. B. (2011). Children's judgments of emotion from conflicting cues in speech: why 6-year-olds are so inflexible. *Child Development*, *82*(5), 1648-1660. doi: 10.1111/j.1467-8624.2011.01624.x

Wiig, E. H., Secord, W. A., & Semel, E. (2013). *Clinical evaluation of language fundamentals: CELF-5*. Pearson.

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: some acoustical correlates. *The Journal of the Acoustical Society of America*, *52*(4B), 1238-1250. doi: 10.1121/1.1913238

Yildirim, S., Bulut, M., Lee, C. M., Kazemzadeh, A., Deng, Z., Lee, S., Narayanan, S., & Busso, C. (2004). "An acoustic study of emotions expressed in speech". Presented at *INTERSPEECH-2044*, 2193-2196.

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*(4), 332-344.