Journal of Experimental Child Psychology 199 (2020) 104907

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

I know that voice! Mothers' voices influence children's perceptions of emotional intensity

Tawni B. Stoop [a,*], Peter M. Moriarty [b], Rachel Wolf [a], Rick O. Gilmore [a], Koraly Perez-Edgar [a], K. Suzanne Scherf [a], Michelle C. Vigeant [b], Pamela M. Cole [a]

[a] Department of Psychology, The Pennsylvania State University, State College, PA 16803, USA
[b] Acoustics Program, College of Engineering, The Pennsylvania State University, State College, PA 16803, USA

ARTICLE INFO

Article history:
Received 1 March 2019
Revised 11 May 2020

Keywords:
Affective prosody
Speaker familiarity
Emotion
Emotional intensity
Acoustic properties
Vocal affect
Children

ABSTRACT

The ability to interpret others' emotions is a critical skill for children's socioemotional functioning. Although research has emphasized facial emotion expressions, children are also constantly required to interpret vocal emotion expressed at or around them by individuals who are both familiar and unfamiliar to them. The current study examined how speaker familiarity, specific emotions, and the acoustic properties that comprise affective prosody influenced children's interpretations of emotional intensity. Participants were 51 7- and 8-year-olds presented with speech stimuli spoken in happy, angry, sad, and nonemotional prosodies by both each child's mother and another child's mother unfamiliar to the target child. Analyses indicated that children rated their own mothers as more intensely emotional compared with the unfamiliar mothers and that this effect was specific to angry and happy prosodies. Furthermore, the acoustic properties predicted children's emotional intensity ratings in different patterns for each emotion. The results are discussed in terms of the significance of the mother's voice in children's development of emotional understanding.

* Corresponding author.
E-mail address: tbs23@psu.edu (T.B. Stoop).

https://doi.org/10.1016/j.jecp.2020.104907
0022-0965/

Introduction

Social information conveyed through speech is complex, and the ability to understand nonlexical information is critical for effective social interaction (Adolphs, 2002). Moreover, decoding emotional cues is an important component of the development of socioemotional competence during childhood (Halberstadt, Denham, & Dunsmore, 2001; McElwain, Halberstadt, & Volling, 2007; Saarni, 1999). However, children's understanding of others' emotions involves more than just comprehending the words that they say. It includes *who* is speaking (i.e., familiar vs. not) and *how* those words are spoken (i.e., affective prosody). The current study examined the relations between children's perceptions of speakers' emotional intensity and three features of speech stimuli: speaker familiarity, the emotion category used by the speaker, and the specific acoustic properties that comprise affective prosody (Banse & Scherer, 1996; Pell, Paulmann, Dara, Alasseri, & Kotz, 2009; Williams & Stevens, 1972).

Speaker familiarity perceptions

On a daily basis, individuals may listen to any number of voices, both familiar and unfamiliar. Whether a voice is familiar to someone can affect several psychological processes. For example, adults are better able to focus on and attend to a familiar voice compared with an unfamiliar voice (Johnsrude et al., 2013; Souza, Gehani, Wright, & McCloy, 2014) and, as a result, perform better on a variety of tasks when attending to that familiar voice compared with the unfamiliar voice (Newman & Evers, 2007; Nygaard, Sommers, & Pisoni, 1994). The beneficial effect of a familiar voice is present even when listeners are not aware that the voice is familiar to them (Holmes, Domingo, & Johnsrude, 2018). School-age children also appear to benefit from familiarity, as suggested by evidence that children's verbal processing of a speaker's words improves after the children become familiar with that speaker, especially when the speaker has an accent (Levi, 2015).

However, familiarity can vary substantially, from repeatedly hearing a voice during a training session to hearing the voice of a close relative. For children, familiarity with their mothers' voices may play a unique role in speech processing. Recognition of mothers' voices has been demonstrated as early as the third trimester (Kisilevsky et al., 2009). Exposure to mothers' voices in utero may explain why newborns prefer their mothers' voices even compared with their fathers' voices (Lee & Kisilevsky, 2013), and infants as young as 7 months can distinguish their own mothers' voices through background noise better than an unfamiliar female voice (Barker & Newman, 2004). Furthermore, neural activity in infants (Imafuku, Hakuno, Uchida-Ota, Yamamoto, & Minagawa, 2014; Naoi et al., 2012) and children (Abrams et al., 2016; Liu et al., 2019) differentiates the sound of their own mothers' voices compared with an unfamiliar female voice. This evidence highlights the salience of their mothers' voices to children, especially when compared with an unfamiliar voice. Regardless of who is speaking, however, children must also interpret the emotional information, the affective prosody, present in speakers' voices.

Affective prosody

Affective prosody adds emotional information that can shape listeners' expectations beyond what is available in semantic cues alone (Paulmann, Titone, & Pell, 2012). As with infants' sensitivity to their mothers' voices, preverbal infants are also sensitive to variations in affective prosody (Fernald, 1993; Walker-Andrews & Grolnick, 1983). Specifically, 5-month-olds' facial expressions affectively match their mothers' approving and disapproving vocalizations (Fernald, 1993). As infants enter their second year of life, they 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 are reluctant to repeat the task after an unfamiliar adult speaks angrily at the research assistant teaching them the task (Repacholi & Meltzoff, 2007). Eye-tracking studies suggest that 4- and 5-year-olds preemptively look at an object (broken vs. intact toy) that aligns with the affective prosody (angry vs. happy) of semantically neutral instructions (i.e., "Look at the . . .") before the target object is identified (i.e., ". . . broken doll") (Berman,

Chambers, & Graham, 2010; Berman, Graham, & Chambers, 2013). Thus, in addition to being sensitive to the familiarity of the speaker, young children detect and react to variations in speakers' affective prosody.

Although these studies provide evidence for the ability to discriminate between valences of affective prosody during the first 5 years of life, other research presents mixed results regarding whether children under 7 years of age, relative to older children, accurately identify specific emotions in the voice based on affective prosody. For example, children between 3 and 6 years of age label the emotion conveyed by affective prosody at or below chance rates of accuracy in speech scripts with nonemotional words, suggesting that they may need semantic content to interpret affective prosody (Aguert, Laval, Lacroix, Gil, & Le Bigot, 2013; Nelson & Russell, 2011). Studies that test the effects of semantic cues that are congruent or incongruent with the speaker's tone of voice indicate that younger children rely on semantic information over affective prosody more than older children do (Friend, 2000; Friend & Bryant, 2000; Morton & Trehub, 2001). In contrast, other studies suggest that 3- to 6-year-olds accurately identify emotions based on affective prosody with unintelligible speech (Baltaxe, 1991; Morton, Trehub, & Zelazo, 2003) or speech that does not include semantic emotional information (Sauter, Panattoni, & Happé, 2013). In addition, children's accuracy in identifying prosody in experimental stimuli improves when children are primed to focus on prosody (Waxer & Morton, 2011), further highlighting their capability to interpret affective prosody. Altogether, these mixed findings suggest that children younger than 7 years might not reliably attend to affective prosody (Friend, 2000, 2003; Friend & Bryant, 2000; Morton & Trehub, 2001) but that they are still sensitive to affective prosody when presented with less complicated congruent stimuli (Baltaxe, 1991; Berman et al., 2010).

Research findings are consistent, however, in showing that children's accuracy at labeling the intended emotion in affective prosodies increases with age (Aguert et al., 2013; Friend, 2000; Morton & Trehub, 2001; Nelson & Russell, 2011; Rothman & Nowicki, 2004; Sauter et al., 2013). Typically developing children of at least 8 years of age perform as well as adults in emotion labeling based on affective prosody (Van Lancker, Cornelius, & Kreiman, 1989). By 9 years of age, they achieve 80% accuracy in labeling emotion based on prosody even if they self-report relying on neutral contextual information (Aguert et al., 2013). In addition, approximately half of 10-year-olds 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 have the ability to distinguish the emotional valence conveyed by prosody (e.g., negative vs. positive, angry vs. happy), and by early school age children can accurately identify emotion in the voice, particularly if presented in ways that are semantically congruent or neutral. We used semantically neutral speech stimuli with 7- and 8-year-olds to examine the extent to which children's perceptions of affective prosody, specifically emotional intensity, may be moderated by speaker familiarity.

The single existing study that also used semantically neutral stimuli found that 7- to 12-year-olds more accurately label the emotion in their mothers' voices relative to an unfamiliar female voice, depending on which emotion is conveyed (Shackman & Pollak, 2005). This work provided support for preferential processing based on who is speaking and the expressed emotion. However, the stimuli involved trials of conflicting facial and vocal emotions, conditions that may have increased the complexity of the recognition task for children (Shackman & Pollak, 2005). The current study aimed to examine how speaker familiarity may influence perceptions of affective prosody without the complication of conflicting facial or verbal information spoken by different groups of women.

Emotional intensity

Much of the evidence for children's understanding of affective prosody focused on labeling accuracy (Friend, 2000; Morton & Trehub, 2001; Sauter et al., 2013). In addition to accuracy, children 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. For example, overhearing parents argue may mean something different to children if their parents are quietly disagreeing compared with yelling loudly.

There is limited research on children's perception of the intensity of emotion cues. Studies using facial stimuli of emotion indicate 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 with their peers (Pollak & Sinha, 2002). For these children, perceiving low-level anger may help them to avoid unwanted consequences.

Just as facial expressions can vary in emotional intensity, affective prosody can also vary in emotional intensity. One study suggests that children's emotion labeling accuracy does not differ during any age period for high-intensity versus low-intensity vocal expressions (Zupan, 2015). However, these results were based on children's responses to a standardized test of emotion recognition; therefore, we do not know whether children interpret the intensity of vocal emotion differently if they are familiar with the speaker. The significance of mothers' emotions for children may lead them to interpret it differently relative to that of an unfamiliar mother. That is, mothers' emotions may have more personal significance than those of unfamiliar mothers, and this significance may lead them to judge emotion in their mothers' voices as more intense. To test that prediction, it is also important to evaluate the acoustic properties of affective prosody that can influence perceptions of emotion intensity. Specifically, the pattern of a range of acoustic properties varies by person, and these properties can influence how others interpret the person's affective prosody.

Acoustic properties

Although psychological factors such as speaker familiarity may affect perceptions of emotion, ample evidence indicates that variations in distinct acoustic properties contribute to adult listeners' abilities to label discrete emotions (Bachorowski & Owren, 2003; Banse & Scherer, 1996; Williams & Stevens, 1972). For example, a higher mean fundamental frequency (F0–pitch; Banse & Scherer, 1996; Pell et al., 2009), higher values on measures of F0 variability (i.e., wider range of pitch values, higher standard deviation, or higher variance; Banse & Scherer, 1996; Pell et al., 2009; Sobin & Alpert, 1999), and a faster speech rate (Breitenstein, Van Lancker, & Daum, 2001) have been associated with the identification of anger. Similarly, happiness is associated with a high mean F0 (Pell et al., 2009), higher values on measures of F0 variability (Pell et al., 2009; Sobin & Alpert, 1999), and a fast rate of speech (Banse & Scherer, 1996; Pell et al., 2009), whereas sadness is associated with a low mean F0, lower values on measures of F0 variability, and a slow speech rate (Banse & Scherer, 1996; Breitenstein et al., 2001; Pell et al., 2009; Sobin & Alpert, 1999).

This evidence suggests that acoustic properties are relevant to adults' perceptions of vocal emotion, but little research has actually examined these acoustic properties in relation to children's perceptions of vocal emotion. Some research has demonstrated that infants prefer child- and infant-directed speech (Schachner & Hannon, 2011), which is typically characterized by higher F0 and higher values on measures of F0 variability compared with typical adult-directed speech (Fernald & Mazzie, 1991). Importantly, infants' preference for infant-directed speech is not dependent on who is being spoken to (Singh, Morgan, & Best, 2002) but instead is influenced by the pitch characteristics of infant-directed speech (Fernald & Kuhl, 1987). Additional work has suggested that 4- and 5-year-olds successfully used pitch characteristics to determine whether a puppet was joyful/happy or sad (Quam & Swingley, 2012). Furthermore, differences in mean F0 have been associated with the speed and accuracy of emotion recognition of joy and anger in children aged 7–17 years (Dmitrieva, Gel'man, Zaitseva, & Orlov, 2008). Speech rate did not have a significant relation with emotion recognition in these children (Dmitrieva et al., 2008), but this study did not examine recognition of sad stimuli, which is the emotion most related to speech rate (Williams & Stevens, 1972).

This limited work focuses exclusively on emotion labeling, leaving open the issue of whether children's perceptions of emotional intensity in the voice are determined by acoustic properties. Studies of adult participants have demonstrated that fluctuations in mean F0 and F0 standard deviation (F0 SD) are associated with ratings of the speaker's level of arousal or intensity of stress (Juslin & Laukka, 2001; Streeter, Macdonald, Apple, Krauss, & Galotti, 1983), but research is needed to confirm that this

relation exists for children. The current study aimed to fill this gap by examining the extent to which the relevant acoustic properties—F0 (pitch) and speech rate—are associated with children's ratings of emotional intensity in their mothers' and unfamiliar mothers' voices.

Finally, it is important to mention the relevance of using unfamiliar mothers as a comparison group. Using a single unfamiliar female individual or trained actors in the unfamiliar condition introduces a potential confound when mothers comprise the familiar condition. Although the existing study of speaker familiarity and affective prosody interactions included mothers as both familiar and unfamiliar stimuli, the groups did not consist of the same women (Shackman & Pollak, 2005). It is possible that the two groups differed in their presentations of emotions or acoustic profiles. Therefore, having the same group of untrained mothers participate as the familiar voice for their own child and as the unfamiliar voice for a different child ensures that recordings and emotion representations are of similar quality for both conditions. Furthermore, hormonal changes during pregnancy alter musculature, including the muscles of the voice (Cassiraga, Castellano, Abasolo, Abin, & Izbizky, 2012). This may explain why mothers' speech, compared with non-mothers' speech, differs in acoustic profiles (Kempe, Schaeffler, & Thoresen, 2010). In our study, children's mothers served as the unfamiliar voices for other children. In this way, differences between the familiar and unfamiliar groups were not attributable to motherhood and variability in producing affective prosody was the same across groups.

The current study

The current study investigated how speaker familiarity (i.e., own mother vs. an unfamiliar mother), emotion category (anger, happiness, and sadness), and the acoustic properties most often associated with these basic emotions (mean F0, F0 SD, and speech rate) may influence children's perceptions of emotional intensity in quasi-natural speech samples. Moving beyond studies of children's accuracy in labeling personally irrelevant experimental stimuli, this study addressed the determinants of children's emotion intensity ratings when the voices presenting the standardized stimuli vary in personal relevance. We created stimuli that use the same scripts of semantically neutral, grammatically correct sentences in the native language (English) of the children. This approach was intended to be more ecologically valid for children than unintelligible, masked, foreign, or nonsense speech, which children do not hear and use in their daily lives. In addition, we varied speaker familiarity, inviting children's own mothers to produce speech stimuli and comparing those stimuli with those of other mothers in the study who were unfamiliar to the children. Furthermore, we focused on children's intensity ratings of affective prosody rather than on accuracy to capture the dynamic nature of emotional information conveyed through human speech.

This study aimed to (a) examine whether children's ratings of emotional intensity differ based on speaker, emotion, or their interaction, (b) determine whether F0, F0 SD, and speech rate influence children's ratings of emotional intensity and whether those influences differ based on which emotion is being rated, and (c) determine whether the influence of speaker familiarity on children's emotional intensity ratings exists even after accounting for the influence of the acoustic properties and whether that influence differs based on the emotion being rated. Based on the literature examining the importance of familiarity on children's emotion recognition (Shackman & Pollak, 2005), we hypothesized that (1) children would rate their own mothers as more intensely emotional than the unfamiliar mothers, especially for anger. Furthermore, we hypothesized that (2) variations of F0, F0 SD, and speech rate would predict children's perceptions of emotional intensity. Guided by prior work with adult participants (Pell et al., 2009), we specifically hypothesized that all three acoustic properties would interact with the emotion being rated such that (a) a high mean F0 and a fast speech rate would influence children's perceptions of anger, (b) a high F0 SD and a fast speech rate would influence children's perceptions of happiness, and (c) a low mean F0 and a slow speech rate would influence children's perceptions of sadness. Finally, we hypothesized that (3) speaker familiarity would influence children's emotional intensity ratings for all emotions even after accounting for the influence of the acoustic parameters and that this influence would be more significant for anger.

Method

Participants

The final sample included 51 children aged 7 years 0 months to 8 years 11 months (25 girls; M_{age} = 8.03 years, *SD* = 0.51) who completed all study visits. Families in a small mid-Atlantic city of the United States were recruited for inclusion in the multi-visit study through a university database, publicly posted flyers, community events, and word of mouth. In addition, 54 mothers (M_{age} = 39.0 9 years, *SD* = 4.78) provided vocal stimuli.

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 the current study were drawn from a larger study involving four data collection points (see online supplementary material). Only select components from the first and fourth visits that are relevant to the current study are described here. At the first visit to the lab, mothers recorded the vocal stimuli that children heard during subsequent visits. During a later visit, children listened to these created speech stimuli to evaluate their perceptions of the speech, including the intensity of the emotions conveyed.

All speech stimuli were presented on a laptop through headphones. To ensure that participants remained on task, a research assistant paused playback and prompted the children immediately following each recording. The task lasted approximately 30 min. Children rated each individual recording in multiple ways. Relevant to the current study, 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 Fig. 1). The current study focused only on children's ratings of angry, happy, and sad emotional intensities and the acoustic properties of the speech stimuli (described next).

Stimuli creation

A total of 54 unique mothers created stimuli for inclusion in the current study. Of these mothers, 6 were not included as unfamiliar stimuli because 3 mothers had strong non-native accents that may have prosodic (Anderson-Hsieh & Koehler, 1988) and speech rate (van Maastricht, Krahmer, & Swerts, 2016) differences compared with native speakers that could negatively affect the speech comprehension of a child unfamiliar with the speaker (Bent & Atagi, 2017; Munro & Derwing, 1995) and 3 mothers were recorded near the end of data collection when no more unfamiliar mother voices could be used. Analyses indicate no significant differences in children's emotional intensity ratings or in acoustic variables between mothers with non-native accents and those with native accents. As a result, 47 unique mothers served as the voice of an unfamiliar mother for another child in the study, and 4 of those mothers were selected twice so that each child heard an unfamiliar mother. 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 unique mothers serving as the voice of a familiar mother. A total of 45 mothers appeared in both groups.

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 the film's actors practiced conveying emotions in their voices, serving as models for the emotions that mothers were asked to produce vocally (i.e., hot anger and joyful happiness). Mothers were then given an opportunity to discuss the task, to recall times when they felt the target emotions, and to practice generating emotion in their voices.

Fig. 1. Visual aid for emotional intensity ratings.

Table 1
Descriptors of the target affects provided to mothers before recording.

Nonemotional	Anger	Sadness	Happiness
Removed	Sharp	Subdued	Melodic
Complacent	Biting	Low energy	Sing song
Apathetic	Harsh	Lethargic	Chirpy
Robotic	Abrupt	Depressed	Sprightly
Flat	Frontal	Exhausted	Chipper
Factual	Raised	Hopeless	Liltingly
News report	Demanding	Breathy	Quick
	Coarse		Hopeful
	Exasperated		

Mothers were next shown the eight semantically nonemotional scripts to review and rehearse before recording began. The research assistant corrected any pronunciation or semantic errors prior to the start of recording and during recording if necessary. Recognizing the novelty of our research, we decided not to use child-directed speech or directly induce emotional reactions from children. Because overhearing parental voices is psychologically meaningful (Repacholi & Meltzoff, 2007), each script was designed and explained to mothers as one side of a telephone conversation with another adult (see supplementary material for the full scripts). Mothers recorded scripts in a randomized order, but with a standardized order of four prosodies: nonemotional, angry, sad, and joyful/happy. As a result, each mother produced 32 stimuli (8 scripts × 4 prosodies).

Recording took place in a sound-isolated room. A graduate student in acoustics 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 noise transfer. Recording equipment included a condenser microphone and shock mount to prevent vibrations from the microphone stand transferring during recording. The sound engineer paused the recording process if extraneous noise was detected in the microphone and also provided feedback to the mother regarding the recording (e.g., that she had moved too close to the microphone) to ensure high recording quality. Each mother completed at least one practice and two usable 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 based on the subjective agreement between the research staff that it was representative of the target emotion and the mother's report of naturalness.

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, 2005) for processing. The stimuli were standardized to all be exactly 10 s in length by inserting or removing silences between utterances. The stimuli were also normalized for loudness (amplitude) due to requirements for another study session. This normalization entailed applying the A-weighted root mean square (RMS) for amplitude to each recording such that within a given recording each utterance was scaled to have the same RMS amplitude value as the utterance in that recording with the lowest RMS amplitude 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 normalization. This normalization 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 (https://www.databrary.org; Cole, Gilmore, Perez-Edgar, & Scherf, 2017).

Children heard all 32 stimuli recorded by their own mothers 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. Stimuli were presented to each child in one of four pseudo-random orders. All four orders began and ended with a nonemotional unfamiliar voice. Within each order, recordings had an equal probability of preceding or following every other type of speaker by emotion recording (see Liu et al., 2019).

Affective prosody variables: Children's ratings and acoustic properties

Intensity ratings

Children heard and provided emotional intensity ratings for 8 nonemotional, 8 angry, 8 happy, and 8 sad recordings each for two speakers. The current study focused on the angry, happy, and sad prosody recordings. As noted, children rated each recording on the perceived intensity of happiness, anger, and sadness. Relevant to the current study, each participant provided emotional intensity ratings for each of 48 recordings to be included in analyses (8 scripts × 3 prosodies × 2 speakers).

Acoustic parameters

Acoustic parameters were curated for each recording using the speech processing software PRAAT (Boersma & Weenink, 2001; https://praat.org) and MATLAB to acquire numeric representations for

each target acoustic parameter (for full details, see Moriarty, Vigeant, Wolf, Gilmore, & Cole, 2018). The sound engineer (acoustics graduate student) created individual code sequences (i.e., scripts) in PRAAT. The target acoustic parameters included F0 mean, F0 SD, and syllabic rate.

F0 mean

This represents the average pitch or pulse rate of the vocal cords, measured in hertz, present in a given recording. For each stimulus, an F0 mean was calculated using an autocorrelation algorithm (Boersma, 1993). This resulted in 48 unique F0 mean values for each participant included in analyses (8 scripts × 3 prosodies × 2 speakers).

F0 Sd

This is a measure of how spread out or variable F0 values (pitches) are in a given recording. F0 SD captures whether a speaker is using a monotone voice or a more dynamic fluctuating voice that might include both very high and very low pitch values. F0 SD was calculated by taking the square root of the variance of the pitches within an utterance and then averaged across the entire recording. This resulted in 48 unique F0 SD values for each participant included in analyses (8 scripts × 3 prosodies × 2 speakers).

Syllabic rate

This is a measure of how quickly or slowly the speaker produces syllables. Syllabic rate was calculated by dividing the number of phonetically counted syllables within an utterance by the spoken duration of that utterance to determine the number of syllables per second. These were established based on the number of syllables in the text of each standardized script and the individual utterance durations created by the participants. This resulted in 48 unique syllabic rate values for each participant (8 scripts × 3 prosodies × 2 speakers).

Data analytic plan

Analyses were completed using the *multilevel* package in R Version 3.4.1 for Windows. The distributions of each variable were visually inspected, and skew and kurtosis values were calculated to assess for violations of normality. Distributions appeared to be normal, and all skew and kurtosis values were less than |1|.

Preliminary analyses

Descriptive statistics and correlations were conducted for all study variables of interest prior to primary data analysis. To determine whether children were accurately identifying the target emotions, *t* tests were run to compare the average intensity ratings of the stimulus target emotion with those of the other two nontarget emotions. Children were deemed to be accurate if the average intensity of the target emotion (e.g., the angry rating for an angry recording) was rated significantly higher than the intensity of the nontarget emotions (e.g., the happy and sad ratings for an angry recording).

Primary analyses

We used a series of two- and three-level models to test hypotheses in repeated measures nested within participants. Given that participants are likely to provide ratings that are more similar to their other ratings than to those of other children, a multilevel model approach provides better estimates by accounting for the nonindependence of multiple scores provided by individual participants (e.g., repeated emotional intensity ratings for multiple stimuli presented to the same child). We did not anticipate that the relations among test variables would differ for each person or for individual recordings; therefore, participant and recording were included as random factors in the relevant models. No covariates were included.[1]

[1] Because research has yielded little evidence of gender differences in vocal affect perception in children (Baltaxe, 1991; Shackman & Pollak, 2005), we did not consider child gender as a factor. Models were also tested with age included as a covariate. The pattern of results was not altered; therefore, the simplified models are presented here.

Speaker and emotion effects

Model 1 tested the hypothesis that children rate their mothers' speech stimuli as more intensely emotional than the unfamiliar mothers' stimuli, and that this may interact with the specific emotion being conveyed. The outcome variable was emotional intensity, formulated as one composite variable that included only the emotional intensity rating that matched the target emotion of the specific recording. The within-participant predictor variables were speaker (dummy coded such that 0 = familiar and 1 = unfamiliar) and emotion (angry, happy, or sad). Post hoc tests were conducted to examine any significant interactions.

Acoustic property and speaker effects

Model 2 was a three-level model that tested the hypothesis that acoustic properties influence children's perceptions of emotional intensity in speech stimuli and that speaker familiarity accounts for intensity ratings even after controlling for those properties. Given that children provided ratings for three separate emotions, these models also examined whether the effects of acoustic properties and speaker interact with the emotion that children were asked to rate. The outcome variable was intensity rating, formulated as one composite variable that included all emotional intensity ratings regardless of whether they matched the target emotion of the specific recording. The within-participant predictor variables were F0 mean, F0 SD, syllabic rate, rated emotion, speaker, and interactions of each acoustic property and speaker with the rated emotion. The predictor acoustic variables were mean-centered to reflect differences from overall average F0 mean, F0 SD, and syllabic rate of all speakers. Ratings were nested within recordings (i.e., each recording was rated for angry, happy, and sad intensity), which were nested within participants. Recording and participant were included as random factors.

Model 3 consisted of a series of parallel two-level models used to understand the interaction effects found in the prior model. For each emotion that participants rated, this model further tested the specific relations among acoustic properties, speaker familiarity, and children's perceptions of emotional intensity in speech stimuli based on the emotion being rated. Separate analyses were completed for each rated emotion. The outcome variables were the separate angry, happy, and sad intensity ratings comprised of the ratings provided for all recordings, including both the target and nontarget emotion recordings. Retention and inclusion of all intensity ratings for even nontarget emotion stimuli is necessary to determine whether the variability in acoustic properties of the stimuli discriminates between the target and nontarget emotion intensity ratings. The within-participant predictor variables were F0 mean, F0 SD, syllabic rate, and speaker. These predictor variables were mean-centered to reflect differences from overall average F0 mean, F0 SD, and syllabic rate of all speakers. Ratings were only nested within participants because only one rating per recording was used for each emotion model.

Results

Descriptive statistics

Means and standard deviations for emotional intensity ratings, F0 mean, F0 SD, and speech rate for the angry, happy, and sad stimuli by speaker 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 indicated that children perceived the emotions as moderately intense. Correlation analyses of all study variables regardless of target emotion category indicated that target variables were significantly associated with each other in the expected directions (Table 3).

Preliminary analysis: Children's accuracy

Prior to the primary analyses, *t* tests were run to determine whether children accurately perceived the emotions conveyed in the recordings. As would be expected if children accurately identified the target emotions, pairwise comparisons revealed that children rated the angry recordings as more

Table 2
Descriptive statistics for acoustic variables 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
Mothers	3.30	0.50	3.21	0.61	2.95	0.65
F0 mean	228.99	32.93	316.97	37.58	183.29	16.85
F0 SD	49.91	11.04	94.19	20.58	22.51	4.40
Speech rate	3.81	0.40	4.21	0.59	3.39	0.41
Unfamiliar mothers	3.02	0.56	2.99	0.76	2.97	0.69
F0 mean	230.22	32.95	319.20	38.97	184.04	17.29
F0 SD	49.18	11.18	93.49	19.98	22.95	4.71
Speech rate	3.85	0.39	4.19	0.60	3.42	0.42

Note. Intensity ratings were on a scale from 1 to 4, with an average score of 3 representing *some* of the target emotion. F0, fundamental frequency; SD, standard deviation.

Table 3
Correlations for children's intensity perceptions and acoustic variables for all recordings.

	Age	Angry rating	Happy rating	Sad rating	F0 mean	F0 SD	Speech rate
Age	–						
Angry rating	.020	–					
Happy rating	−.014	−.400***	–				
Sad rating	.079***	−.231***	−.375***	–			
F0 mean	–	.131***	.637***	−.476***	–		
F0 SD	–	−.115***	.623***	−.513***	.882***	–	
Speech rate	–	.016	.298***	−.290***	.395***	.371***	–

Note. F0, fundamental frequency; SD, standard deviation.
*** $p < .001$.

intensely angry than intensely happy or sad (*p*s < .001) (Fig. 2). They also rated happy recordings as more intensely happy than intensely angry or sad, and they rated sad recordings as more intensely sad than intensely angry or happy (all *p*s < .001).

Speaker and emotion effects

Model 1 tested for differences in children's emotional intensity ratings of the target emotion based on speaker, emotion category, and their interaction (Table 4). Consistent with our hypothesis, results suggested a main effect of speaker such that children rated their own mothers as more intensely emotional overall than unfamiliar mothers ($b = -0.461$, $SE = 0.098$, $p < .001$). Results also identified a significant main effect of emotion ($b = -0.176$, $SE = 0.032$, $p < .001$). Follow-up analyses using planned contrasts determined that children perceived sad recordings as less intensely emotional than angry and happy recordings ($b = -0.101$, $SE = 0.018$, $p < .001$), but they did not rate the intensity of angry and happy recordings differently ($b = 0.050$, $SE = 0.032$, $p = .118$).

These main effects were qualified by a significant interaction (Fig. 3) between speaker familiarity and emotion category in predicting children's intensity ratings ($b = 0.146$, $SE = 0.045$, $p = .001$). Specifically, children rated their own mothers as more intensely emotional, but only for angry ($b = -0.290$, $SE = 0.064$, $p < .001$) and happy ($b = -0.220$, $SE = 0.064$, $p < .001$) recordings. Emotional intensity ratings did not differ by speaker for sad recordings ($b = 0.003$, $SE = 0.064$, $p = .969$).

Acoustic property and speaker effects

The next model tested the relations of acoustic properties, speaker, the rated emotion, and their interaction to emotional intensity across all ratings provided by children (Model 2). Model 3 then

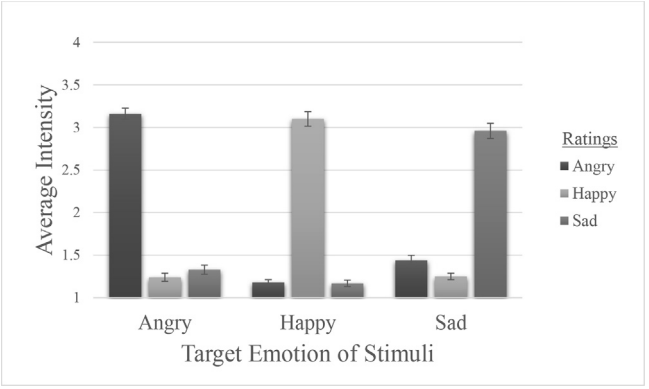

Fig. 2. Children's intensity ratings for each asked emotion by target emotion of the stimuli.

Table 4
Results of multilevel model testing for difference in children's target emotional intensity ratings by emotion and by speaker.

	Model 1	
	b (*SE*)	*p*
Fixed effects		
Intercept	3.509 (0.092)	
Speaker	**−0.462 (0.098)**	**<.001**
Emotion	**−0.176 (0.032)**	**<.001**
Interaction	**0.146 (0.045)**	**.001**
Random effects		
Intercept	0.439 (0.905)	–

Note. Bold values are significant.

Fig. 3. Interaction of children's emotional intensity ratings by speaker and emotion. Average intensity ratings represent only the target emotion intensity rating. ***$p < .001$.

expanded the significant interaction effects by testing these relations for each rated emotion separately to examine the specific acoustic profiles of each emotion.

Model 2 tested the relations among the acoustic properties, speaker familiarity, the emotion being rated (i.e., anger, happiness, or sadness), and their interaction in children's perceptions of emotional intensity in audio recordings. Results indicated a main effect of F0 SD and speech rate such that a wider F0 SD (b = 0.011, SE = 0.002, p < .001) and a faster speech rate (b = 0.304, SE = 0.053, p < .001) were associated with perceptions of higher emotional intensity. Results also suggested no main effect of F0 mean on children's perceptions of emotional intensity (Table 5). Results found a main effect of the rated emotion (b = −0.056, SE = 0.023, p = .013) and were consistent with results found in Model 1 such that children rated sadness as less intensely emotional than anger and happiness. Results also indicated a main effect of speaker familiarity after controlling for the effects of acoustic properties and rated emotion such that children rated their mothers' voices as more intensely emotional than the unfamiliar mothers' voices (b = −0.211, SE = 0.069, p = .002).

As hypothesized, interaction analyses indicated that the effects of F0 SD and speech rate were moderated by the specific emotion that children were asked to rate, suggesting that these properties may have differential relations for predicting the intensity of anger, happiness, and sadness. Furthermore, results indicated that the effect of speaker familiarity was moderated by the emotion that participants were asked to rate (Table 5) such that the relevance of speaker familiarity differed for anger, happiness, and sadness. Next, Model 3 further examined the nature of this interaction.

Model 3 further examined the interaction effects by testing the relations among the acoustic properties, speaker familiarity, and children's perceptions of anger, happiness, and sadness in audio recordings separately based on the emotion that children were asked to rate. As hypothesized, variations in F0 mean, F0 SD, and speech rate each had a unique relation to the specific emotion category being rated (Table 6). However, the specific relations were not all as hypothesized. Children's perceptions of anger intensity were associated with a lower than average F0 mean (b = −0.004, SE = 0.001, p < .001) and a faster than average speech rate (b = 0.151, SE = 0.038, p < .001), whereas children's perceptions of anger intensity were not associated with F0 SD. Children's perceptions of happiness intensity were associated with higher than average F0 mean (b = 0.007, SE = 0.000, p < .001) and higher F0 SD (b = 0.008, SE = 0.001, p < .001), but perceptions of happiness intensity were not associated with speech rate. Finally, as hypothesized, children's perceptions of sadness intensity were associated with a lower than average F0 mean (b = −0.001, SE = 0.001, p = .020), a lower F0 SD (b = −0.012, SE = 0.001, p < .001), and a slower than average speech rate (b = −0.196, SE = 0.030, p < .001). Results also indi-

Table 5
Results of multilevel models testing for relations among acoustic properties, speaker familiarity, and children's perceptions of emotional intensity.

	Model 2	
	b (SE)	p
Fixed effects		
Intercept	2.001 (0.054)	
F0 mean	−0.001 (0.001)	.541
F0 SD	**0.011 (0.002)**	**<.001**
Speech rate	**0.304 (0.053)**	**<.001**
Rated emotion	**−0.056 (0.023)**	**.013**
Speaker	**−0.211 (0.069)**	**.002**
F0 Mean × Rated Emotion	0.001 (0.001)	.159
F0 SD × Rated Emotion	**−0.006 (0.001)**	**<.001**
Speech Rate × Rated Emotion	**−0.150 (0.024)**	**<.001**
Speaker × Rated Emotion	**0.086 (0.032)**	**.007**
Random effects		
Participant	0.174	–
Recording	0 (1.105) (Bänziger & Scherer, 2005; Scherer, Koivumaki, & Rosenthal, 1972)	

Note. Bold values are significant. F0, fundamental frequency; SD, standard deviation.

Table 6
Results of multilevel models examining interactions among acoustic properties, speaker familiarity, and children's perceptions of emotional intensity based on the rated emotion.

	Model 3	
	b (*SE*)	*p*
Angry		
Fixed effects		
Intercept	1.960 (0.046)	
F0 mean	**−0.004 (0.001)**	**<.001**
F0 SD	0.001 (0.001)	.524
Speech rate	**0.151 (0.038)**	**<.001**
Speaker	**−0.103 (0.046)**	**.025**
Random effects		
Intercept	0.232 (1.12)	–
Happy		
Fixed effects		
Intercept	1.861 (0.045)	
F0 mean	**0.007 (0.000)**	**<.001**
F0 SD	**0.008 (0.001)**	**<.001**
Speech rate	0.049 (0.029)	.092
Speaker	**−0.083 (0.034)**	**.016**
Random effects		
Intercept	0.272 (0.840)	–
Sad		
Fixed effects		
Intercept	1.848 (0.045)	
F0 mean	**−0.001 (0.001)**	**.020**
F0 SD	**−0.012 (0.001)**	**<.001**
Speech rate	**−0.196 (0.030)**	**<.001**
Speaker	0.070 (0.036)	.054
Random effects		
Intercept	0.263 (0.885)	–

Note. Bold values are significant. F0, fundamental frequency; SD, standard deviation.

cated that speaker familiarity was significantly associated with children's perceptions of emotional intensity even after accounting for the influence of acoustic properties, but only for anger (b = −0.102, SE = 0.046, p = .025) and happiness (b = 0.083, SE = 0.034, p = .016). Speaker familiarity was not associated with emotional intensity ratings after accounting for acoustic properties for sadness (b = 0.070, SE = 0.036, p = .054).

Discussion

This study tested the hypothesis that children would rate their own mothers' voices as more intensely emotional than unfamiliar mothers' voices and that this relation may depend on the emotion that each mother expressed. Furthermore, we tested whether the acoustic properties that comprise angry, happy, and sad speech predicted children's ratings of emotional intensity and whether this relation depended on the emotion that children needed to rate. Finally, we tested the hypothesis that speaker familiarity would influence children's intensity ratings even after accounting for the influence of acoustic properties.

In support of our first hypothesis, children rated their own mothers as more intensely emotional compared with unfamiliar mothers. As predicted, this effect of speaker familiarity depended on which emotion was conveyed. Specifically, children perceived their own mothers as more intensely angry and more intensely happy than unfamiliar mothers, but they did not perceive a difference for sadness. Notably, this speaker familiarity effect goes beyond any influences due to pregnancy (Cassiraga et al., 2012) or motherhood (Kempe et al., 2010) given that the same group of mothers provided both the

familiar and unfamiliar voices. As a result, differences in children's ratings of emotional intensity for anger and happiness for *their own* mothers compared with those for other mothers were perceptual.

These results are similar to findings from Shackman and Pollak (2005), who found that children more accurately identified anger and happiness in their own mothers' voices compared with unfamiliar women's voices. Notably, that study included a large sample of maltreated children for whom emotion type may have had unique significance. The current findings are based on typically developing children with no identified risk, suggesting that this familiarity effect with mothers may be generalizable. Furthermore, our findings are in line with evidence that infants and children attend more to positive and negative emotions expressed by a familiar person, particularly their mothers (Arsalidou, Barbeau, Bayless, & Taylor, 2010; Kahana-Kalman & Walker-Andrews, 2001; Montague & Walker-Andrews, 2002; Shackman & Pollak, 2005; Todd, Evans, Morris, Lewis, & Taylor, 2011).

Children in the current study may have perceived their mothers as more intensely emotional than unfamiliar mothers because the children were inclined to attend more carefully to their mothers' voices. Given that 7- and 8-year-olds have spent a significant amount of time with their mothers throughout their young lives, their mothers' emotional states may have more direct and immediate consequences for the children. As a result, children may be motivated to be more sensitive to their mothers' emotional states, as opposed to those of unfamiliar mothers, in order to better anticipate potential outcomes in their daily lives, particularly when their mothers are angry or happy.

That explanation depends on outcome relevance and learned patterns, but we cannot rule out the influences of extrinsic cognitive processes. For example, children may intentionally attend to their mothers' voices upon realizing that one voice is their mother and another voice is completely unfamiliar. This type of top-down influence has been demonstrated for perception of sentences in unintelligible altered speech samples (Remez, Rubin, Pisoni, & Carrell, 1981) and for recognition of a previously unrecognized song in altered musical patterns once participants have learned of its identity (Deutsch, 1972). In these instances, knowing what they would hear influenced participants' ability to perceive it. Children in the current study were informed that they would hear voices including their mothers' voices, and this knowledge may have influenced our findings.

Alternatively, it is possible that children rated their mothers as more intensely emotional than unfamiliar mothers because children are better able to discern differences in their mothers' emotional states. For example, throughout their daily lives, children experience the entire range of their mothers' emotions, from elation to passive amusement and from disappointment to hot anger. Their lack of experience with an unfamiliar voice may limit their sensitivity to or interest in the differences between the unfamiliar mother's intense anger and mild anger. Furthermore, all mothers were instructed to provide a clear example of how they usually expressed each emotion; because the recording of their vocal emotion was nonetheless an unusual circumstance, they may have communicated emotion, or at least some emotions, more intensely than they usually do at home, which may have contributed to children's perceptions of higher intensity. Ratings for the unfamiliar mothers' voices could not have been influenced by children's preexisting perceptions given that we confirmed that those voices were unfamiliar to children. Therefore, children were unable to compare the recordings of unfamiliar mothers as different than usual.

Instead of children having expertise with their mothers' emotions, it is also possible that children's mothers are their exemplars for emotional expressions. Research examining word and speech–sound learning has demonstrated that children are better able to distinguish between changes in similar familiar words (Fennell & Werker, 2003) and melodies (Creel, 2019) but struggle to distinguish between those same changes in unfamiliar words (Stager & Werker, 1997) and melodies (Creel, 2019). Creel (2019) argued that children learn by complete patterns, or exemplar samples, rather than by specific subcomponents (e.g., syllables, rising vs. falling pitch). The current findings align with this interpretation; children may have learned emotions and emotional variation as entire representations of pitch, pitch standard deviation, and speech rate fluctuations that their mothers present. Unfamiliar mothers' presentations, even if their acoustic properties (pitch, pitch standard deviation, and speech rate independently) are the same as the children's own mothers, may differ holistically from the exemplar model that children have learned. This difference may then explain why children perceive their own mothers' emotions as more intense than unfamiliar mothers' emotions.

These possibilities collectively highlight the significance of mothers in children's daily lives and how these daily experiences influence the development of emotional awareness. Future work should, however, aim to determine which of the potential explanations is at the root of the familiarity effect. This work may consider the inclusion of positive or negative consequences for performance that could increase the relevance of unfamiliar mothers' voices. If the familiarity effect were to disappear in this paradigm, it would provide more support for outcome relevance over prior knowledge/expertise of the speakers themselves. Future work may also consider testing for top-down influences by manipulating what children are told about the speech stimuli. For example, children could be told that their mothers are very angry or a little angry for the same speech sample, and resulting differences in their perceptions of emotional intensity would support top-down effects.

As predicted, wider F0 SD and faster speech rate were associated with higher ratings of emotional intensity, but F0 mean was not associated with overall intensity ratings. This pattern of results suggests that although all three acoustic properties are associated with emotion identification, F0 SD and speech rate may be most informative for overall emotional intensity. Prior literature has demonstrated that these acoustic properties differentially relate to anger, happiness, and sadness; therefore, examining their interaction with the emotion being rated is crucial. As hypothesized, the relations of F0 SD and speech rate with emotional intensity were moderated by the emotion that children were asked to rate. Although the interaction between F0 mean and rated emotion was not significant, we included it in the follow-up analyses for completeness given the differential relations for each emotion demonstrated in prior literature. Further examination revealed that F0 mean, F0 SD, and speech rate each had differential influences on children's perceptions of emotional intensity for angry, happy, and sad stimuli. Specifically for F0 mean, these differences were small, which may explain the nonsignificant interaction. Broadly, the results are consistent with existing literature; prior research indicates that young children use the average pitch (mean F0) to identify joy and anger in the voice (Dmitrieva et al., 2008), and different emotion categories have different acoustic profiles that include F0 mean, F0 SD, and speech rate (Banse & Scherer, 1996; Williams & Stevens, 1972). Furthermore, existing adult literature suggests that F0 mean and F0 SD influence listeners' perceptions of the intensity of a speaker's stress (Juslin & Laukka, 2001; Streeter et al., 1983). The exact patterns of relations, however, were not all as expected based on the adult literature.

In the current study, higher perceived intensity of sadness was associated with lower average pitch, a smaller standard deviation of pitch, and a slower speech rate, which is consistent with existing adult literature (Banse & Scherer, 1996; Breitenstein et al., 2001; Pell et al., 2009; Sobin & Alpert, 1999). Higher perceived intensity of happiness was associated with higher average pitch and higher standard deviations of pitches used. Speech rate was not associated with perceived intensity of happiness even though some existing literature suggests that a faster speech rate contributes to identification of happiness (Banse & Scherer, 1996; Pell et al., 2009). Higher perceived intensity of anger was associated with a lower average pitch and a faster speech rate, whereas pitch standard deviation was not associated. This contrasts with findings that higher average pitch and higher values on measures of pitch variability are associated with adults' identification of anger (Banse & Scherer, 1996; Pell et al., 2009; Sobin & Alpert, 1999). There are a few key differences between those published studies and the current study that may contribute to these discrepancies.

First, the current study used mothers with little to no acting experience, whereas some of the previous studies used trained actors or standardized stimuli for all vocal samples (Banse & Scherer, 1996; Breitenstein et al., 2001; Pell et al., 2009). Actor training alters speakers' F0 mean and range (Walzak, McCabe, Madill, & Sheard, 2008). However, there still appear 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 the current study's acoustic profiles do not exactly match the existing literature.

More likely, the current study emphasized emotional intensity ratings rather than emotion labeling (Banse & Scherer, 1996; Pell et al., 2009). Changes in emotional intensity are dependent on and potentially more sensitive to the variability that occurs in acoustic properties of speech samples. In contrast, emotion labeling requires that speech samples generally follow a pattern to still be included in a specific emotion category. The patterns of acoustic variation needed to distinguish between levels of inten-

sity within a category may understandably be different from those needed to distinguish between different emotion categories.

In support of our third hypothesis, the effect of speaker familiarity emerged even after controlling for any influences of the acoustic properties. Interaction analyses revealed that this result was specific for perceptions of angry and happy intensity. In other words, speaker familiarity played an important role in children's perceptions of emotional intensity ratings for anger and happiness regardless of the effects of acoustic properties. Broadly, these findings are consistent with top-down influences on emotion perception. Specifically, prior research has demonstrated that participants will hear a sentence when told what to listen for even if those speech samples have been acoustically altered to eliminate intelligible speech cues (Remez et al., 1981). In the current study, participants' perceptions of emotional intensity were altered beyond the base acoustic information provided to them as a result of familiarity.

That this effect existed only for angry and happy stimuli 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 (Todd, Lewis, Meusel, & Zelazo, 2008). This work reported that children exhibited a significantly longer processing latency following mothers' angry faces but not following an unfamiliar woman's face, suggesting that children spent more time in processing their mothers' angry faces. In addition, vocal emotion research has suggested that both typically developing and physically abused children focus more on angry vocal expressions in their mothers' voices than on angry vocal expressions of an unfamiliar woman (Shackman & Pollak, 2005). However, prior work has not highlighted the salience of happiness in familiar speakers compared with unfamiliar speakers or compared with other emotions. Speakers often have similar levels of arousal or energy when conveying anger and happiness (Banse & Scherer, 1996; Jallais & Gilet, 2010; Yildirim et al., 2004), which may explain why the familiarity effect is consistent for both emotions.

Overall, these data underscore the fact that emotional understanding of anger and happiness may be less about the physical properties of voices in general and more about *who* is angry or happy. Furthermore, it is unclear whether this effect may be situational. For example, children in the current study listened to contextually nonemotional adult-directed speech such that their mothers or unfamiliar mothers were speaking over the phone to an unknown adult. For many children in a novel research situation, an unfamiliar woman's angry or happy voice *not* directed at them might not be meaningful. Their own mothers' angry voice, even directed at another adult, is more 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 adult may still affect children's actions—for example, if a child is being taught to do something that makes that unfamiliar woman angry (Repacholi & Meltzoff, 2007) and the child's own mother is not speaking.

It is unclear whether these findings extend to emotional speech directed at a child—either a same-age other child or the target child. Using stimuli without a clear target listener, physically abused children more readily identify vocal anger in their mothers' voices than in unfamiliar women's voices (Shackman & Pollak, 2005). For children exposed to physical abuse, emotion-laden child-directed speech from a familiar voice may also be judged as more intense. On the other hand, children who regularly hear anger in their mothers' voices but do not experience negative consequences as a result might not perceive differences in the intensity of their mothers' and unfamiliar mothers' voices when speech is directed at them or at other children. The current findings suggest the need for additional research on different forms of emotional speech, including adult-directed versus child-directed speech, to clarify the conditions that govern how speaker familiarity influences children's interpretation of emotional intensity.

Finally, analyses of the effect of emotion on intensity ratings, regardless of speaker familiarity, suggested a difference in children's intensity ratings for sad recordings compared with angry and happy recordings. Children perceived sad recordings as less intensely sad than angry recordings were intensely angry or happy recordings were intensely happy. One possible interpretation of this finding could be that the children may have had more difficulty in understanding and identifying variations in sadness. This would be consistent with literature suggesting that, at least for facial emotion, children's abilities to identify sadness increase with age, but children are consistently able to identify anger from

an early age (Herba et al., 2008). Angry and happy recordings are usually more easily identifiable; thus, young children may have had an easier time in perceiving intensity differences in these emotions. However, sadness is also more acoustically different from anger and happiness than they are from each other; it is likely that children accurately perceived acoustic differences, which were reflected in their intensity ratings. More specifically, vocal anger and happiness require higher levels of arousal and energy, both of which are acoustic indicators of intensity and easier to mimic compared with sadness, which has the lowest levels (Banse & Scherer, 1996; Jallais & Gilet, 2010; Yildirim et al., 2004). Given that their intensity ratings are consistent with this directional acoustic difference, our findings that acoustic properties influence children's perceptions of emotional intensity are supported.

Both explanations (i.e., children struggle to identify sadness and are sensitive to acoustic differences in emotions), taken together, further support the significance of speaker familiarity demonstrated by our earlier findings. Children are able to accurately perceive acoustic differences in intensity, as highlighted by their lower intensity ratings for sadness compared with anger and happiness. However, speaker familiarity causes a psychological effect on children's perceptions of intensity that is not present in the acoustic profiles, especially given that the two groups of speakers consist of the same women and are acoustically equivalent. In the current study, the psychological effect is demonstrated in the emotions that children most easily identify (i.e., anger and happiness) and to which they are more sensitive. It makes sense that children would not experience this psychological difference for sadness given that they generally have difficulty in identifying sadness at all (Herba et al., 2008). It will be important for future research to examine whether older children demonstrate an effect of speaker familiarity on perceptions of intensity of sadness at a point when they are more able to identify sadness. If the familiarity effect emerges, it would provide further support for the significance of speaker familiarity on emotion understanding and emotional development.

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 continue to examine other acoustic properties such as energy to establish whether other properties of the voice may also contribute to perception of emotional intensity specifically rather than emotion labeling. As previously noted, children perceived sadness at lower levels of intensity, which tends to differ from the other emotions in levels of mean energy (Banse & Scherer, 1996; Jallais & Gilet, 2010; Yildirim et al., 2004).

Second, the current study required that children make explicit judgments of differences in emotional intensity. We did not control the speakers' intensity levels, which leads to two additional questions: (a) whether children were accurately perceiving emotional intensity and (b) whether children may implicitly understand and perceive differences in intensity more accurately than they are able to do so explicitly. Future work should consider (a) controlling the variations in emotional intensity to determine whether children accurately perceive changes in emotional intensity and (b) examining markers such as neural reactivity to determine whether there are differences in children's implicit and explicit sensitivity to emotional intensity.

Third, one strength of the study design involved children rating each recording on the intensity of three separate emotions, which acknowledges the dynamic and nuanced nature of emotion expression. However, we did not directly compare children's simultaneous ratings of angry, happy, and sad intensity within the same recording. Future work should consider comparing these types of ratings to test children's sensitivity to nuanced and mixed emotions (e.g., a bittersweet feeling) conveyed in the voice to determine whether children focus on or perceive one emotion over others in these complex expressions. Current research often examines discrete, clear examples of individual emotions and not the more natural expressions of emotion that may be less clear-cut and more nuanced.

Fourth, this study examined children in a limited age range (7- and 8-year-olds), which in some studies is less accurate than older children in identifying emotions in the voice. Longitudinal studies are lacking in this line of research, but they would be useful in elucidating whether and when acoustic properties influence children's perceptions of emotional intensity, as well as the factors that con-

tribute to this developing ability. Furthermore, prior evidence demonstrated the salience of their mothers' voices, but the extent of these effects and how the level of salience may change over time remain unknown. Speaker familiarity was an important factor in emotional intensity perception for this group of 7- and 8-year-olds, but future research should examine whether this effect remains or is stronger at younger or older ages.

Finally, the current study focused on only three emotions—anger, happiness, and sadness. Prior literature suggests that infants and children attune to other maternal emotions (e.g., fear) to influence their behavior (Gerull & Rapee, 2002; Mumme et al., 1996; Sorce, Emde, Campos, & Klinnert, 1985). Future research should consider testing whether speaker familiarity plays a role in children's perceptions of the intensity of other emotions.

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 F0, F0 SD, 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 and happy prosody. These results also suggested that children use acoustic properties in speech to determine emotional intensity. However, differences in perceived intensity are still influenced by more psychological factors related to speaker familiarity regardless of the influence of acoustic parameters.

Acknowledgments

This material is based on work supported by the National Institute of Mental Health under Grant 1R21MH104547-01A1 awarded to Pamela M. Cole. 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.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jecp.2020.104907.

References

Abrams, D. A., Chen, T., Odriozola, P., Cheng, K. M., Baker, A. E., Padmanabhan, A., ... Menon, V. (2016). Neural circuits underlying mother's voice perception predict social communication abilities in children. *Proceedings of the National Academy of Sciences of the United States of America, 113*, 6295–6300.

Adobe Systems. (2015). Adobe Audition [computer program]. San Jose, CA: Author. Retrieved from https://www.adobe.com/Audition.

Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology, 12*, 169–177.

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.

Anderson-Hsieh, J., & Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning, 38*, 561–613.

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*, 47–51.

Bachorowski, J., & Owren, M. J. (2003). Sounds of emotion. *Annals of the New York Academy of Sciences, 1000*, 244–265.

Baltaxe, C. A. M. (1991). Vocal communication of affect and its perception in three- to four-year-old children. *Perceptual and Motor Skills, 72*, 1187–1202.

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

Bänziger, T., & Scherer, K. R. (2005). The role of intonation in emotional expressions. *Speech Communication, 46*, 252–267.

Barker, B. A., & Newman, R. S. (2004). Listen to your mother! The role of talker familiarity in infant streaming. *Cognition, 94*, B45–B53.

Bent, T., & Atagi, E. (2017). Perception of nonnative-accented sentences by 5- to 8-year-olds and adults: The role of phonological processing skills. *Language and Speech, 60*, 110–122.

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*, 87–99.

Berman, J. M. J., Graham, S. A., & Chambers, C. G. (2013). Contextual influences on children's use of vocal affect cues during referential interpretation. *Quarterly Journal of Experimental Psychology, 66*, 705–726.
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*, 97–110.
Boersma, P., & Weenink, D. (2001). PRAAT (Version 4.0) [computer software]. Amsterdam: University of Amsterdam. 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*, 57–79.
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*, 584–586.
Cole, P. M., Gilmore, R. O., Perez-Edgar, K., & Scherf, K. S. (2017). Children's neural processing of the emotional environment (PEEP-II).. *Databrary*. https://doi.org/10.17910/b7.339.
Creel, S. C. (2019). The familiar-melody advantage in auditory perceptual development: Parallels between spoken language acquisition and general auditory perception. *Attention, Perception, & Psychophysics, 81*, 948–957.
Deutsch, D. (1972). Octave generalization and tune recognition. *Perception & Psychophysics, 11*, 411–412.
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*, 527–531.
Fennell, C. T., & Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and Speech, 46*, 245–264.
Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development, 64*, 657–674.
Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development, 10*, 279–293.
Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology, 27*, 209–221.
Friend, M. (2000). Developmental changes in sensitivity to vocal paralanguage. *Developmental Science, 3*, 148–162.
Friend, M. (2003). What should I do? Behavior regulation by language and paralanguage in early childhood. *Journal of Cognition and Development, 4*, 161–183.
Friend, M., & Bryant, J. B. (2000). A developmental lexical bias in the interpretation of discrepant messages. *Merrill-Palmer Quarterly, 46*, 342–369.
Gerull, F. C., & Rapee, R. M. (2002). Mother knows best: Effects of maternal modelling on the acquisition of fear and avoidance behaviour in toddlers. *Behaviour Research and Therapy, 40*, 279–287.
Halberstadt, A. G., Denham, S. A., & Dunsmore, J. C. (2001). Affective social competence. *Social Development, 10*, 79–119.
Herba, C. M., Benson, P., Landau, S., Russell, T., Goodwin, C., Lemche, E., ... Phillips, M. (2008). Impact of familiarity upon children's developing facial expression recognition. *Journal of Child Psychology and Psychiatry, 49*, 201–210.
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*, 1098–1106.
Holmes, E., Domingo, Y., & Johnsrude, I. S. (2018). Familiar voices are more intelligible, even if they are not recognized as familiar. *Psychological Science, 29*, 1575–1583.
Imafuku, M., Hakuno, Y., Uchida-Ota, M., Yamamoto, J., & Minagawa, Y. (2014). "Mom called me!" Behavioral and prefrontal responses of infants to self-names spoken by their mothers. *NeuroImage, 103*, 476–484.
Jallais, C., & Gilet, A. (2010). Inducing changes in arousal and valence: Comparison of two mood induction procedures. *Behavior Research Methods, 42*, 318–325.
Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., & Carlyon, R. P. (2013). Swinging at a cocktail party: Voice familiarity aids speech perception in the presence of a competing voice. *Psychological Science, 24*, 1994–2004.
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.
Juslin, P. N., & Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion, 1*, 381–412.
Kahana-Kalman, R., & Walker-Andrews, A. S. (2001). The role of person familiarity in young infants' perception of emotional expressions. *Child Development, 72*, 352–369.
Kempe, V., Schaeffler, S., & Thoresen, J. C. (2010). Prosodic disambiguation in child-directed speech. *Journal of Memory and Language, 62*, 204–225.
Kisilevsky, B. S., Hains, S. M. J., Brown, C. A., Lee, C. T., Cowperthwaite, B., Stutzman, S. S., ... Wang, Z. (2009). Fetal sensitivity to properties of maternal speech and language. *Infant Behavior and Development, 32*, 59–71.
Lee, G. Y., & Kisilevsky, B. S. (2013). Fetuses respond to father's voice but prefer mother's voice after birth. *Developmental Psychobiology, 56*, 1–11.
Levi, S. V. (2015). Talker familiarity and spoken word recognition in school-age children. *Journal of Child Language, 42*, 843–872.
The Mathworks (2005). *MATLAB (Version 7.0) [computer software]*. Natick, MA: Author.
Liu, P., Cole, P. M., Gilmore, R. O., Perez-Edgar, K., Vigeant, M. C., Moriarty, P., & Scherf, K. S. (2019). Young children's neural processing of their mother's voice: An fMRI study.. *Neuropsychologia, 22*, 11–19. https://doi.org/10.1016/j.neuropsychologia.2018.12.003.
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*, 1407–1425.
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*, 1339–1352.
Moriarty, P. M., Vigeant, M., Wolf, R., Gilmore, R. O., & Cole, P. M. (2018). Creation and characterization of an emotional speech database.. *The Journal of the Acoustical Society of America, 143*(3), 1869. https://doi.org/10.1121/1.5036133.
Morton, J. B., & Trehub, S. E. (2001). Children's understanding of emotion in speech. *Child Development, 72*, 834–843.
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*, 1857–1868.

Mumme, D. L., Fernald, A., & Herrera, C. (1996). Infants' responses to facial and vocal emotional signals in a social referencing paradigm. *Child Development, 67*, 3219–3237.
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*, 289–306.
Naoi, N., Minagawa-Kawai, Y., Kobayashi, A., Takeuchi, K., Nakamura, K., Yamamoto, J., & Kojima, S. (2012). Cerebral responses to infant-directed speech and the effect of talker familiarity. *NeuroImage, 59*, 1735–1744.
Nelson, N. L., & Russell, J. A. (2011). Preschoolers' use of dynamic facial, bodily, and vocal cues to emotion. *Journal of Experimental Child Psychology, 110*, 52–61.
Newman, R. S., & Evers, S. (2007). The effect of talker familiarity on stream segregation. *Journal of Phonetics, 35*, 85–103.
Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5*, 42–46.
Paulmann, S., Titone, D., & Pell, M. D. (2012). How emotional prosody guides your way: Evidence from eye movements. *Speech Communication, 54*, 92–107.
Pell, M. D., Paulmann, S., Dara, C., Alasseri, A., & Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics, 37*, 417–435.
Pixar Animation Studios. (2015). Inside Out behind the scenes footage—Amy Poehler, Bill Hader, Mindy Kaling, Lewis Black [online video]. Emeryville, CA: Author. 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*, 784–791.
Quam, C., & Swingley, D. (2012). Development in children's interpretation of pitch cues to emotions. *Child Development, 83*, 236–250.
Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 212*, 947–949.
Repacholi, B. M., & Meltzoff, A. N. (2007). Emotional eavesdropping: Infants selectively respond to indirect emotional signals. *Child Development, 78*, 503–521.
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*, 67–92.
Saarni, C. (1999). *The development of emotional competence*. New York: Guilford.
Sauter, D. A., Panattoni, C., & Happé, F. (2013). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology, 31*, 97–113.
Schachner, A., & Hannon, E. E. (2011). Infant-directed speech drives social preferences in 5-month-old infants. *Developmental Psychology, 47*, 19–25.
Scherer, K. R., Koivumaki, J., & Rosenthal, R. (1972). Minimal cues in the vocal communication of affect: Judging emotions from content-masked speech. *Journal of Psycholinguistic Research, 1*, 269–285.
Shackman, J. E., & Pollak, S. D. (2005). Experiential influences on multimodal perception of emotion. *Child Development, 76*, 1116–1126.
Singh, L., Morgan, J. L., & Best, C. T. (2002). Infants' listening preferences: Baby talk or happy talk?. *Infancy, 3*, 365–394.
Sobin, C., & Alpert, M. (1999). Emotion in speech: The acoustic attributes of fear, anger, sadness, and joy. *Journal of Psycholinguistic Research, 28*, 347–365.
Sorce, J. F., Emde, R. N., Campos, J., & Klinnert, M. D. (1985). Maternal emotional signaling: Its effects on the visual cliff behavior of 1-year-olds. *Developmental Psychology, 21*, 195–200.
Souza, P., Gehani, N., Wright, R., & McCloy, D. (2014). The advantage of knowing the talker. *Journal of the American Academy of Audiology, 24*, 689–700.
Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature, 388*, 381–382.
Streeter, L. A., Macdonald, N. H., Apple, W., Krauss, R. M., & Galotti, K. M. (1983). Acoustic and perceptual indicators of emotional stress. *Journal of the Acoustical Society of America, 73*, 1354–1360.
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*, 12–23.
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*, 595–613.
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*, 261–269.
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.
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.
Walker-Andrews, A. S., & Grolnick, W. (1983). Discrimination of vocal expressions by young infants. *Infant Behavior and Development, 6*, 491–498.
Walzak, P., McCabe, P., Madill, C., & Sheard, C. (2008). Acoustic changes in student actors' voices after 12 months of training. *Journal of Voice, 22*, 300–313.
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*, 1648–1660.
Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *Journal of the Acoustical Society of America, 52*, 1238–1250.
Yildirim, S., Bulut, M., Lee, C. M., Kazemzadeh, A., Deng, Z., Lee, S., . . ., Busso, C. (2004, October). An acoustic study of emotions expressed in speech. In Proceedings of InterSpeech (pp. 2193–2196), Jeju Island, Korea.
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.