



Contents lists available at ScienceDirect

## Journal of Experimental Child Psychology

journal homepage: [www.elsevier.com/locate/jecp](http://www.elsevier.com/locate/jecp)



# Intersensory redundancy promotes infant detection of prosody in infant-directed speech

Lorraine E. Bahrick<sup>a,\*</sup>, Myriah E. McNew<sup>a</sup>, Shannon M. Pruden<sup>a</sup>,  
Irina Castellanos<sup>b</sup>

<sup>a</sup> Department of Psychology, Florida International University, Miami, FL 33199, USA

<sup>b</sup> Department of Otolaryngology—Head & Neck Surgery, The Ohio State University, Columbus, OH 43212, USA

### ARTICLE INFO

#### Article history:

#### Keywords:

Infant prosody detection  
Infant attention  
Intersensory redundancy  
Infant-directed speech  
Audiovisual synchrony  
Audiovisual speech

### ABSTRACT

Prosody, or the intonation contours of speech, conveys emotion and intention to the listener and provides infants with an early basis for detecting meaning in speech. Infant-directed speech (IDS) is characterized by exaggerated prosody, slower tempo, and elongated pauses, all amodal properties detectable across the face and voice. Although speech is an audiovisual event, it has been studied primarily as a unimodal auditory stream without the synchronized dynamic face of the speaker. According to the intersensory redundancy hypothesis, redundancy across the senses facilitates perceptual learning of amodal information, including prosody. We predicted that young infants who are still learning to discriminate and categorize prosodic information would detect prosodic changes better in the presence of intersensory redundancy (i.e., synchronous audiovisual speech) than in its absence (i.e., unimodal auditory or asynchronous audiovisual speech). To test this hypothesis, 72 4-month-old infants were habituated to recordings of women reciting passages in IDS with prosody conveying either approval or prohibition and then were tested with recordings of a novel passage with either a *change* or *no change* in prosody. Infants who received bimodal synchronous stimulation exhibited significant visual recovery to the novel passage with a change in prosody, but not to a novel passage with no change in prosody. Infants in the unimodal auditory and bimodal asynchronous conditions did not exhibit visual recovery in either condition. Results support the hypothesis that intersensory redundancy facilitates detection and abstraction of invariant

\* Corresponding author.

E-mail addresses: [bahrick@fiu.edu](mailto:bahrick@fiu.edu) (L.E. Bahrick), [mmcnew@fiu.edu](mailto:mmcnew@fiu.edu) (M.E. McNew).

prosody across changes in linguistic content and likely serves as an early foundation for the detection of meaning in fluent speech.

© 2019 Published by Elsevier Inc.

## Introduction

To break into language learning, infants are faced with the challenge of parsing what they hear from a continuous speech stream into discriminable units (i.e., words). Infant-directed speech (IDS), also known as motherese, provides the naïve perceiver valuable information in the form of frequent and elongated pauses, slower tempo, pitch changes (i.e., higher pitch and wider pitch range), and more prosodic repetition (Fernald, 1984, 1989; Ladd, Silverman, Tolkmitt, Bergmann, & Scherer, 1985; Newport, Gleitman, & Gleitman, 1977). These exaggerated prosodic features (or intonation contours) characterizing IDS provide opportunities for infants to begin to parse the speech stream and perceive meaning in speech (Morgan, 1996; Nazzi, Kemler Nelson, Jusczyk, & Jusczyk, 2000; Soderstrom, 2007; Spinelli, Fasolo, & Mesman, 2017). In IDS, emotional expressions are also exaggerated, making it easier to accurately detect affective information in the face (Juslin & Laukka, 2001; Ladd et al., 1985). Furthermore, caregivers use IDS to elicit infant attention, communicate meaning, and maintain social interactions (Bryant & Barrett, 2007; Fernald, 1984; Spinelli et al., 2017; Trainor, Austin, & Desjardins, 2000). Decades of research indicate that infants benefit significantly from adults' use of IDS. These studies demonstrate not only that infants prefer to listen to IDS over adult-directed speech (ADS; e.g., Cooper & Aslin, 1990; Fernald, 1985) but also that the unique prosodic patterns found in IDS promote better outcomes during infancy and childhood, including attention, language learning, and discrimination of emotions or affective information (Saint-Georges et al., 2013; Santarcangelo & Dyer, 1988; Spinelli et al., 2017; Werker & McLeod, 1989). The affective intent of speech is linked to specific acoustic profiles (e.g., happiness is characterized by a slower rate of speech and wider expansion of pitch range; anger is characterized by a short sharp tone and narrow pitch range; Fernald, 1993; Juslin & Laukka, 2003; Sakkalou & Gattis, 2012; Scherer, 1986, 2003). The coordination of affective and acoustic information is exaggerated in IDS. Prosodies conveying approval and praise (e.g., "Good baby!") are characterized by exaggerated rise-fall pitch contours and sustained volume intensity, whereas prohibition and warning prosodies (e.g., "No, don't touch!") are characterized by low pitch, high intensity, and short staccato contours (Fernald & Kuhl, 1987; Fernald, 1989). Adults (both with and without experience with infants) are able to identify the communicative intent of a speaker using only prosodic information in bids of approval and prohibition (Fernald, 1989). These results highlight the important role discrimination of prosodic characteristics plays in conveying communicative intent and affect to the listener.

Given the importance of perceiving prosody for learning language, as well as the consistent use of IDS within and across cultures by caregivers and non-caregiving adults (Fernald et al., 1989), the current study examined the conditions that promote infant detection of changes in prosody. Prosody and affect discrimination have typically been studied as *vocal* expressions (e.g., Moore, Spence, & Katz, 1997; Soderstrom, 2007; Spence & Moore, 2003; Trainor et al., 2000). However, speech is a multisensory event, providing coordinated and synchronized changes across the face, voice, and gesture for amodal properties specifying prosodic information (Bahrick & Lickliter, 2002, 2014; Gibson, 1969). Amodal information is information that can be conveyed across more than one sense modality, including timing (such as rhythm, tempo, and duration) and intensity patterns that specify affect and communicative intent in audiovisual speech. Similarly, emotion has been characterized as a multicomponent process across feeling, physiology, and expression, with expression reflected in the face, voice, and gesture (Johnstone & Scherer, 2000; Scherer, 2003). Thus, prosody signifying approval versus prohibition is available not only as a vocal signal but also through correlated changes in the movements of the face (e.g., rhythm, tempo, duration, and intensity changes) as well as through the

rising and falling pitch of the voice synchronized with rising and falling movements of the cheeks, forehead, and eyebrows.

#### *Audiovisual synchrony facilitates infant detection of changes in prosody*

The *intersensory redundancy hypothesis* posits that information presented in temporal synchrony and redundantly across sensory modalities (e.g., auditory, visual) facilitates attention and perceptual learning about amodal information, particularly in young infants (Bahrick & Lickliter, 2000, 2002, 2014; Bahrick, Flom, & Lickliter, 2002). Prosodic patterns characterizing approval versus prohibition are conveyed by synchronized changes in the tempo, rhythm, and duration of speech, amodal properties detectable across both the face and voice. Research has demonstrated that young infants are skilled at detecting these amodal temporal properties. For example, infants detect changes in the tempo (Bahrick et al., 2002) and rhythm (Bahrick & Lickliter, 2000) of an audiovisual event more easily and earlier in development when the audible and visible information is presented together in synchrony (e.g., a toy hammer tapping a particular rhythm) rather than when it is presented in just one sense modality alone (auditory or visual) or out of synchrony. Thus, we expected that the face–voice synchrony provided by audiovisual speech would facilitate the early detection of prosodic changes.

#### *Importance of infant detection of prosody*

The characteristic prosody found in IDS has several important contributions to infant attention and learning (Colombo, Frick, Ryther, Coldren, & Mitchell, 1995). Researchers have posited that the function of IDS is threefold: to regulate infant attention, to highlight the structure of language in adult speech for language-learning children, and to help infants interpret incoming affective information from others (Cooper, Abraham, Berman, & Staska, 1997; Fernald, 1989; Grieser & Kuhl, 1988; Singh, Morgan, & Best, 2002). In support of these claims, research examining the benefits of the prosody found in IDS has shown that it (a) aids in the promotion or maintenance of infant attention to faces, voices, and eye gaze (Kaplan, Goldstein, Huckleby, & Cooper, 1995; Saint-Georges et al., 2013; Senju & Csibra, 2008; Spinelli et al., 2017) as well as to language (Fernald & Mazzie, 1991; Ramírez-Esparza, García-Sierra, & Kuhl, 2014; Werker & McLeod, 1989); (b) highlights the syntactic or grammatical structure of language (Fernald & Mazzie, 1991; Ramírez-Esparza et al., 2014; Werker & McLeod, 1989) and the lexical meaning of individual words (Golinkoff & Alioto, 1995; Ma, Golinkoff, Houston, & Hirsh-Pasek, 2011; Song, Demuth, & Morgan, 2010), consequently leading to improved language outcomes (Ramírez-Esparza et al., 2014); and (c) helps infants to interpret affective information and discriminate between emotions conveyed in faces and voices (Fernald, 1989).

Some have argued that it is the emotion or emotional expressiveness of IDS prosody that sets it apart from ADS (Singh et al., 2002; Trainor et al., 2000). Trainor et al. (2000) examined acoustic samples of both IDS and ADS and contended that reported differences between IDS and ADS emerge as a result of the differences in emotional expression conveyed in each type of speech registered, with more widespread and varied emotion conveyed in IDS and more inhibited expression of emotion conveyed in ADS. Singh et al. (2002) also suggested that the greater affect in IDS as compared with ADS contributes to infant preferences for IDS over ADS. In their study, they held affective information constant while presenting unimodal IDS and ADS samples and found that 6-month-olds do not show a significant preference for either speech register. These findings highlight the unique and important role that affective information in speech plays in infant attention to IDS. They also point to the need for further research examining how infants detect changes in prosody that conveys affective information such as that conveying approval and prohibition.

#### *Development of infant detection of prosody*

Even young infants are keen perceivers of affect and prosody. Infants show early preferences for prosodic contours that contain positive affect, such as approval and comfort, over those that contain

negative affect, such as prohibition (Fernald, 1993; Papoušek, Bornstein, Nuzzo, Papoušek, & Symmes, 1990). By 4 months of age, infants show preferences for IDS conveying approval over IDS conveying disapproval (Papoušek et al., 1990). Infants also show more positive affect themselves (e.g., smiling) for IDS conveying approval when compared with IDS conveying prohibition (Fernald, 1993). This was the case across 5-month-olds learning multiple languages, suggesting a cross-cultural preference for positive affect in IDS. Spence and colleagues (Moore et al., 1997; Spence & Moore, 2003) examined in two separate publications 6-month-olds' ability to discriminate and categorize affective prosody. Using an infant-controlled familiarization–test paradigm, infants were familiarized with a set of IDS utterances in prosodies specifying either approval or comfort and then were presented with a novel instance of either the *familiar* prosody (control group; e.g., if familiarized with comfort utterances, they received a novel comfort utterance) or the *novel* prosody (experimental group; e.g., if familiarized with comfort utterances, they received an approval utterance). In one set of studies, Moore et al. (1997) found that 6-month-olds from the experimental group could form categories of affective prosody when they used low-pass filtered utterances, in which the linguistic content of the utterances had been masked but the prosodic features of the utterances, such as pitch, rhythm, and intensity, were preserved and attenuated. In a follow-up study, Spence and Moore (2003) showed that 6-month-olds in the experimental group, but not 4-month-olds, could discriminate and categorize approval and comfort utterances even when utterances were unfiltered, containing the full range of frequencies that naturally occur in IDS. These studies show that by 5 or 6 months of age, infants detect differences in affective prosody, including approval and prohibition. However, one commonality across the studies reviewed above is that infants were presented with prosody in IDS while viewing either no visual information or static nonaffective visual information such as a checkerboard pattern. Thus, these studies leave open the question of whether at a younger age infants could detect changes in prosody in audiovisual speech if the speech samples were accompanied by the dynamically moving face of the speaker, providing intersensory redundancy, as is typical in the natural environment.

*Multimodal presentation* has been shown to promote infant detection of affect in faces and voices. Caron, Caron, and MacLean (1988) found that 5-month-olds, but not 4-month-olds, could discriminate the emotional expressions of happiness and sadness when presented in a multimodal context. A study by Walker-Andrews and Grolnick (1983) suggests that 5-month-old infants can reliably discriminate between happy and sad affective utterances but appear to do so only in conditions where facial expressions accompany the vocal expressions. Walker-Andrews and Lennon (1991) also found evidence that 5-month-olds can discriminate changes in the vocal expressions of happy and angry affects. Infants detected a change in vocal affect, but only when the soundtrack was accompanied by a face and not when it was accompanied by a checkerboard. These studies raise the question of what exactly it is about multimodal presentations that facilitate infant detection of affect and prosody.

#### *Intersensory redundancy as a basis for facilitating detection of affect*

Research generated by the intersensory redundancy hypothesis indicates that it is the redundancy across synchronous facial and vocal information that facilitates detection of affect. By comparing detection of affect in the presence of intersensory redundancy (synchronous audiovisual speech) versus the absence of intersensory redundancy (asynchronous audiovisual speech; unimodal auditory speech; unimodal visual speech), Flom and Bahrick (2007) demonstrated the critical role of intersensory redundancy in bootstrapping infant detection of affect in audiovisual speech. At 4 months of age infants discriminated affective information (e.g., happy, sad, angry) in synchronous audiovisual speech, at 5 months they discriminated the affect in auditory speech, and only by 7 months did they discriminate the affect in unimodal visual speech. Affect was not discriminated in asynchronous audiovisual speech, demonstrating that temporal synchrony between the audio and visual information was necessary for infant discrimination. Thus, similar to findings from studies of infant detection of rhythm (Bahrick & Lickliter, 2000) and tempo (Bahrick et al., 2002), intersensory redundancy provided by audiovisual synchrony is necessary for promoting discrimination early in infancy. Thus, we predicted that this should also be true for infant detection of prosodic information at 4 months of age.

*The current study: does intersensory redundancy promote infant detection of prosody specifying approval versus prohibition?*

The current study was designed to assess whether intersensory redundancy facilitates infants' ability to abstract prosodic information specifying approval versus prohibition. We examined whether infants detected a change in prosody, from approval to prohibition or from prohibition to approval, in conditions where intersensory redundancy (i.e., temporal synchrony) was present versus absent. Intersensory redundancy is present during synchronous audiovisual speech but is absent in asynchronous audiovisual speech and unimodal auditory speech. Using an infant-controlled habituation paradigm, we asked under which of these three conditions 4-month-olds could detect a change in prosody. If intersensory redundancy bootstraps early detection of prosodic changes, we predicted that infants would detect these changes in the presence, but not in the absence, of intersensory redundancy.

Furthermore, in each condition we assessed whether infants could generalize prosodic information to a new speech passage, similar to the design used by [Spence and Moore \(2003\)](#). If so, this would provide data to suggest that infants could abstract invariant information specifying prosodic information across changes in speech passages. Infants were randomly assigned to condition (bimodal synchronous, unimodal auditory, or bimodal asynchronous) and prosody change test type (change or no change). They were habituated to a passage conveying either approval or prohibition and then were tested with a novel passage conveying either a new (change) prosody or the familiar (no change) prosody. We predicted that 4-month-olds would detect the invariant prosodic information across multiple passages and discriminate a change in prosody when given bimodal synchronous stimulation but not when given unimodal auditory or bimodal asynchronous stimulation. Furthermore, the asynchronous audiovisual condition provides the same amount and type of stimulation as the synchronous audiovisual condition and, thus, serves as a control for a number of possible alternative interpretations of differences between the two conditions, including differential arousal effects of the two prosodies. Thus, any differences between the synchronous and asynchronous conditions could be attributed to intersensory redundancy (i.e., audiovisual temporal synchrony).

## Method

### *Participants*

A total of 72 4-month-old infants ( $M = 125.61$  days,  $SD = 3.96$ ) participated in the current study. Of these, 38 were male and 34 were female. All infants were delivered full-term ( $>37$  gestational weeks) without complications and had Apgar scores of 9 or greater. Regarding race/ethnicity, 59 infants were Hispanic, 9 were non-Hispanic White, and 4 were non-Hispanic Black. Families were either English-Spanish bilingual or monolingual English speakers. An additional 18 infants were tested but were excluded from analyses due to experimenter error ( $n = 3$ ), fussiness ( $n = 3$ ), failure to meet the fatigue criterion ( $n = 10$ ; see "Procedure" section for details), or failure to habituate ( $n = 2$ ).

### *Stimuli*

The stimulus events were eight color videotaped recordings depicting one of two women reciting one of two passages in one of two prosodic patterns. Woman A was light-skinned with shoulder-length light brown hair, and Woman B was olive skinned with long dark brown hair. In each video, the woman's face and shoulders were recorded against a uniform blue background. Both passages consisted of three phrases that were recited in English IDS. Passage 1 consisted of the phrases "Look at you," "Come over here by me," and "Where's the baby going?" Passage 2 consisted of the phrases "You did this," "Gentle with the baby," and "Whose doggy is that?" Each passage contained approximately the same number of syllables (15 and 14, respectively) and was spoken in two prosodic patterns specifying approval and prohibition. The women's facial expressions were naturalistic and appropriate to the prosodic patterns conveyed (i.e., positive/happy for approval and negative/angry

for prohibition), similar to infants' experience in their natural environments. Descriptive information for the acoustic properties of our stimuli are presented in [Table 1](#). Consistent with descriptions in the literature, passages specifying approval were characterized by higher and more variable pitch, wider pitch range, and slower rates of speech than passages specifying prohibition.

Adults ( $N = 15$  college students) also rated the affective quality of each synchronous audiovisual stimulus (prosody of approval and prohibition for Women A and B reciting Passages 1 and 2) as positive, neutral, or negative. All adult raters accurately categorized each of the eight videos for each of the two actresses and each passage (i.e., positive for approval and negative for prohibition).

Each of the eight recordings was edited to create three versions, one for each condition: (a) bimodal synchronous, (b) unimodal auditory, and (c) bimodal asynchronous. The bimodal synchronous recordings depicted the dynamically moving woman producing natural synchronous audiovisual IDS. The unimodal auditory recordings depicted the static nonmoving face of the woman in three different poses presented with the auditory recordings used for the bimodal synchronous condition. The bimodal asynchronous recordings depicted the same recordings used in the bimodal synchronous condition, but the auditory and visual information was temporally misaligned (out of synchrony). This was achieved by delaying the soundtrack by 3 s with respect to the video so that one phrase was heard while a different phrase was seen. Thus, the degree of asynchrony was outside the infants' temporal integration window (see [Lewkowicz, 1996](#), for details). In this design, the bimodal asynchronous condition serves as a control for the bimodal synchronous condition given that both conditions offer the same face and voice events, with the same types and total amounts of stimulation, but differ in whether or not they provide intersensory redundancy (i.e., audiovisual temporal synchrony). Thus, differences found between these two conditions would reflect detection of intersensory redundancy (synchrony) while controlling for any differences in arousal, preference for one prosodic pattern over another, or low-level auditory or visual information (e.g., facial expression or facial or vocal feature). Finally, a recording of a green and white plastic toy turtle whose arms and legs spun and produced a whirring sound was used as a control display.

### *Procedure*

Infants were tested to determine whether they could detect a change in passage with or without a change in prosody specifying approval versus prohibition following redundant bimodal audiovisual stimulation compared with nonredundant unimodal auditory stimulation and nonredundant bimodal audiovisual stimulation. Infants were tested using an infant-controlled habituation paradigm (see [Horowitz, 1975](#); [Horowitz, Paden, Bhana, & Self, 1972](#)), which allows individual infants to control the length of each trial with their looking behavior. Infants were randomly assigned to either the bimodal synchronous condition ( $n = 24$ ), unimodal auditory condition ( $n = 24$ ), or bimodal asynchronous condition ( $n = 24$ ). The prosody infants received for habituation (approval vs. prohibition), the woman they received for habituation (Woman A vs. Woman B), and the prosody test change type (change vs. no change) were counterbalanced between infants. Two women were used as stimulus events to ensure that findings were not specific to a particular face/voice but similar across two women. See [Table 2](#) for an overview of the experimental design and the counterbalancing of prosody, woman, and passage within each condition.

Each infant was habituated to one of the two women (Woman A or Woman B) reciting one of the two passages (Passage 1 or Passage 2) in one of the two prosodic patterns (specifying approval or prohibition). Within each of the three conditions (i.e., bimodal synchronous, unimodal auditory, and bimodal asynchronous), half of the infants were randomly assigned to the no prosody change test and the remaining half to the prosody change test. In the no prosody change test condition, infants received test trials depicting the familiar woman reciting a novel passage in the familiar prosody. For example, if an infant was habituated to Woman A reciting Passage 1 in the approval prosody, the infant would then receive test trials with Woman A reciting Passage 2 in the approval prosody. In the prosody change condition, infants received test trials depicting the familiar woman reciting a novel passage in the novel prosody. For example, if an infant was habituated to Woman A reciting Passage 1 in the prosody specifying approval, the infant would then receive test trials with Woman A



**Table 1**

Acoustic characteristics of passages conveying approval and prohibition.

Prosody	Mean pitch (Hz)	SD pitch (Hz)	Pitch range (Hz)	Mean duration (s)	SD duration (s)	Syllables per second
Approval	304.00	36.73	351.12	1.43	0.30	3.47
Prohibition	281.93	31.77	303.33	0.805	0.18	5.99

Note. Each measure is averaged across 24 passages (two repetitions of 6 passages for each of two actresses) for each prosodic pattern.

**Table 2**

Procedure for each condition and prosody change test type.

Condition	Stimulus event	Habituation (N = 72)	Test: No prosody change (n = 36)	Test: Prosody change (n = 36)
Bimodal synchronous (n = 24)	Woman	A	A	A
		B	B	B
	Passage	1	2	2
		2	1	1
Unimodal auditory (n = 24)	Prosody	Approval	Approval	Prohibition
		Prohibition	Prohibition	Approval
	Woman	A	A	A
		B	B	B
Bimodal asynchronous (n = 24)	Passage	1	2	2
		2	1	1
	Prosody	Approval	Approval	Prohibition
		Prohibition	Prohibition	Approval

Note. During test trials, all participants received the familiar woman and a novel passage relative to habituation with either a change in prosody or no change in prosody.

reciting Passage 2 in the prosody specifying prohibition. Visual recovery to the test trials was assessed to determine whether infants detected a change from habituation to test.

In the bimodal synchronous condition, all trials (habituation and test) were presented with audio-visual face–voice synchrony. The unimodal auditory condition trials consisted of the same soundtrack used in the bimodal synchronous condition. To maintain infant attention, they were accompanied by three different static images of the face of the corresponding woman. In the bimodal asynchronous condition, all trials consisted of the same soundtrack and visual recordings used in the synchronous condition but played out of temporal synchrony such that the phrase the infant heard did not align with the phrase the infant saw.

Several aspects of the design ensured that any visual recovery would reflect detection of prosodic information rather than simple discrimination of low-level featural differences. All infants received a novel passage during the test phase (relative to habituation) in order to assess abstraction of invariant prosody across changes in linguistic content (see Gibson, 1969, for more information about invariant detection). Using this design (rather than one with a change in prosody only) ensured that any visual recovery found was unlikely to be based on detection of changes in low-level information (specific to the vocal inflections or visual changes accompanying a particular phrase) but instead on detection of higher-order information common to both passages. Furthermore, finding parallel results across two different actresses (Woman A and Woman B) also would make it unlikely that findings were based on low-level information characterizing changes in the appearance or voice of a specific actress.

The habituation procedure (similar to that of Bahrick et al., 2002; Bahrick & Lickliter, 2000) began with a control trial depicting a toy turtle (attention getter) and proceeded with four mandatory habituation trials. Each trial began when the infant visually fixated the monitor and terminated when the infant looked away for 1.5 s or when 60 s had elapsed. Habituation trials were administered until infants' visual attention decreased to the habituation criterion (50% reduction in visual attention relative to mean looking on the first two habituation [baseline] trials) on two consecutive trials. Infants then received two post-habituation trials identical to the habituation trials. Post-habituation trials were administered to reduce the likelihood of chance habituation. Post-habituation trial looking times were required to meet the same habituation criterion as habituation trials.

Following the habituation and post-habituation trials, infants were administered two test trials depicting a new passage with a change or no change in prosody. All infants were shown the same woman they saw during habituation; however, she was reciting a novel passage. For example, if an infant saw Woman A reciting Passage 1 during habituation, the infant would see Woman A reciting Passage 2 during test. Half of the infants in each condition received test trials depicting the same prosody (approval or prohibition) they had received during habituation, and the other half received test trials depicting a change in prosody relative to habituation. Infants were then administered a final control trial depicting the toy turtle to assess possible fatigue.

Infants' ability to detect the change in prosody was inferred by *visual recovery*, an increase in looking time to test trials depicting a novel prosody (but not to test trials depicting the familiar prosody) relative to looking time during post-habituation trials. To ensure that infants were not fatigued, initial and final control trials were compared. Infants whose visual fixation to the final control trial was less than 35% of their visual fixation to the first trial were considered fatigued and were excluded from analyses ( $n = 10$  [bimodal synchronous  $n = 3$ , unimodal auditory  $n = 3$ , bimodal asynchronous  $n = 4$ ]; see Bahrick et al., 2002, for details).

## Results

### Planned analyses

To determine whether infants discriminated the prosody change, we calculated visual recovery scores by subtracting mean visual fixation time on post-habituation trials from mean visual fixation time on test trials in each condition. Visual recovery scores significantly greater than zero indicate discrimination. Our primary hypothesis—that at 4 months of age infants would require intersensory redundancy to detect prosody change—was tested in two ways: first, by looking at differences among groups in visual recovery and, second, by comparing each group's visual recovery scores with the chance value of zero. In evaluating group differences in an analysis of variance (ANOVA), we expected an interaction effect such that infants would show significantly greater visual recovery to a new passage spoken with a change in prosody than to the new passage with no change in prosody if they could detect invariant prosodic information. When evaluating prosody detection using visual recovery against chance performance, we expected that infants in the bimodal synchronous condition, but not in the bimodal asynchronous or unimodal auditory condition, would show significant visual recovery to a new passage spoken with a change in prosody, but not to the new passage spoken with no change in prosody.

### Primary analyses

To address the main hypothesis, we conducted a  $3 \times 2$  ANOVA on visual recovery scores with condition (bimodal synchronous, unimodal auditory, or bimodal asynchronous) and prosody change test type (change or no change) as between-participants factors. Results indicated a significant main effect of prosody change,  $F(1, 66) = 5.848$ ,  $p = .018$ ,  $\eta_p^2 = .081$ . Participants who received a change in prosody showed greater average visual recovery ( $M = 3.633$ ,  $SD = 7.970$ ) than participants who received no prosody change ( $M = -0.084$ ,  $SD = 5.324$ ). Consistent with our predictions, this main effect was qualified by a significant condition by prosody change test type interaction,  $F(2, 66) = 3.495$ ,  $p = .036$ ,  $\eta_p^2 = .096$ .



As expected, independent-samples *t* tests (corrected for familywise error where appropriate throughout)<sup>1</sup> indicated that infants who were provided with passages containing a change in prosody showed significantly greater visual recovery than infants who were provided with passages containing no change in prosody in the bimodal synchronous condition,  $t(22) = -3.380$ ,  $p = .003$ ,  $d = 1.38$  (significant when adjusted for multiple comparisons,  $p = .05/2 = .025$ ), but not in the unimodal auditory condition,  $t(22) = -0.295$ ,  $p = .771$ ,  $d = 0.12$ , or the bimodal asynchronous condition,  $t(22) = -0.362$ ,  $p = .721$ ,  $d = 0.15$  (see Fig. 1). In support of our hypothesis, results indicated that 4-month-olds show significantly greater visual recovery to a novel passage with a change in prosody than with no change in prosody following redundant bimodal synchronous stimulation, but not following unimodal auditory or bimodal asynchronous stimulation.<sup>2</sup>

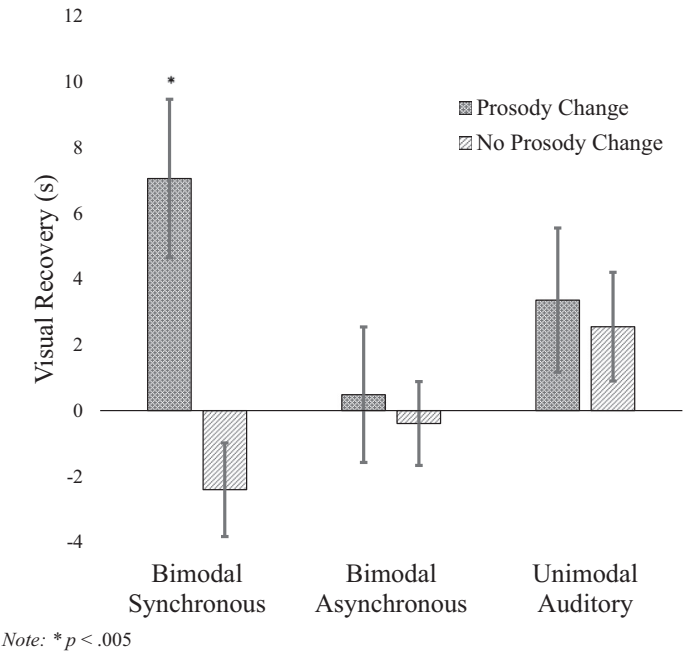
Second, we compared visual recovery scores in each condition alone against the chance value of zero using single-sample *t* tests to assess evidence of discriminating prosody. Results (see Fig. 1) indicated that participants in the bimodal synchronous condition showed significant visual recovery following a change in prosody,  $t(11) = 2.93$ ,  $p = .014$ ,  $d = 0.85$  (significant when adjusted for multiple comparisons,  $p = .05/2 = .025$ ), but not following no prosody change,  $t(11) = -1.691$ ,  $p = .12$ ,  $d = 0.49$ . Visual recovery scores were not significantly different from chance in the bimodal asynchronous condition or unimodal auditory condition following either a change in prosody or no change in prosody ( $ps > .15$ ). These results support our hypothesis and indicate that audiovisual redundancy available in bimodal synchronous stimulation facilitates discrimination of a prosody change (from approval to prohibition or vice versa) across a change in passage in 4-month-old infants. Furthermore, at 4 months infants show no evidence of detecting a change in passage alone (without a change in prosody) under any condition.

### Secondary analyses

Secondary analyses were conducted to determine whether infants' looking behaviors during habituation varied across conditions. A multivariate analysis of variance (MANOVA) was conducted with condition (bimodal synchronous, unimodal auditory, or bimodal asynchronous) as the between-participants factor and mean baseline looking, mean number of habituation trials, mean post-habituation looking, and mean processing time (total number of seconds looking during habituation) as dependent measures (see Table 3). Results indicated significant main effects of condition on mean baseline looking,  $F(2, 69) = 5.795$ ,  $p = .005$ ,  $\eta_p^2 = .144$ , and total processing time,  $F(2, 69) = 6.358$ ,  $p = .003$ ,  $\eta_p^2 = .156$ . Planned pairwise comparisons indicated that infants in the bimodal asynchronous condition displayed significantly greater mean looking during baseline ( $M = 52.47$ ,  $SD = 13.13$ ) than infants in the unimodal auditory condition ( $M = 37.22$ ,  $SD = 16.97$ ),  $p = .004$ . This remained significant when controlling for multiple comparisons,  $p = .05/3 = .017$ . Further, infants in the bimodal asynchronous condition also displayed significantly greater baseline looking than infants in the bimodal synchronous condition ( $M = 41.66$ ,  $SD = 17.43$ ),  $p = .05$ . This *p*-value, however, did not meet the criterion for significance when controlling for multiple comparisons ( $p = .05/2 = .025$ ). In contrast, infants' baseline looking did not differ between the bimodal synchronous and unimodal auditory conditions

<sup>1</sup> For all relevant analyses, planned *a priori* comparisons were conducted using a modified, multistage Bonferroni procedure to control the familywise error rate for multiple comparisons (Holm, 1979; Jaccard & Guilamo-Ramos, 2002). For example, to reach significance in cases where three comparisons were made, the comparison with the lowest *p* value was required to pass a criterion of  $.05/3 = .017$ , the comparison of the next lowest *p* value was required to pass  $.05/2 = .025$ , and the last comparison was required to pass  $.05$ . This method was applied to all cases involving two or more comparisons.

<sup>2</sup> We also conducted another analysis consistent with our analytic approach to take into account individual variation in baseline looking: a repeated-measures ANOVA with trial type (mean baseline looking, mean post-habituation looking, or mean test looking) as a within-participants factor and condition (bimodal synchronous, unimodal auditory, or bimodal asynchronous) as a between-participants factor. Using this approach with participants who received a change in prosody from habituation to test, we found a significant main effect of trial type,  $F(2, 66) = 136.43$ ,  $p < .001$ , whereby infants showed longer looking on baseline trials than on post-habituation and test trials ( $ps < .001$ ). Consistent with our findings using visual recovery, this was qualified by a significant interaction of condition and trial type,  $F(4, 66) = 3.48$ ,  $p = .012$ , whereby infants showed longer looking to test trials than to post-habituation trials in the bimodal synchronous condition ( $p = .014$ ) but not in the unimodal auditory or bimodal asynchronous condition ( $ps > .10$ ). These results complement those reported above while taking into account individual differences in initial looking levels (baseline).



**Fig. 1.** Mean visual recovery to a novel passage with a change in prosody versus a novel passage with no change in prosody as a function of condition. Error bars represent standard errors. \* $p < .005$ .

**Table 3**

Means (and standard deviations) for visual fixation times (in seconds) and number of habituation trials administered as a function of condition.

Condition	Baseline looking	Trials to Habituation	Post-habituation looking	Processing time	Test looking
Bimodal synchronous	41.66 (17.43)	8.04 (2.10)	7.36 (4.96)	258.10 (133.96)	9.69 (8.53)
Unimodal auditory	37.22 (16.97)	7.63 (2.32)	6.23 (4.49)	189.59 (102.31)	9.18 (6.91)
Bimodal asynchronous	52.47 (13.13)	8.33 (2.43)	8.63 (5.38)	318.04 (135.35)	8.68 (5.74)

Note. Baseline is the first two habituation trials. Post-habituation is the two no-change trials following habituation. Processing time is the total time (in seconds) spent fixating the habituation events.

( $p = .602$ ). Despite greater baseline looking time in the asynchronous condition, infants did not exhibit discrimination of prosody. Planned pairwise comparisons also indicated that infants in the bimodal asynchronous condition displayed significantly more total processing time ( $M = 318.04$ ,  $SD = 135.53$ ) than infants in the unimodal auditory condition ( $M = 189.59$ ,  $SD = 102.31$ ),  $p = .002$  (significant when adjusted for multiple comparisons [ $p = .05/3 = .017$ ]). However, total processing time in the bimodal synchronous condition ( $M = 258.10$ ,  $SD = 133.96$ ) did not differ from that in the bimodal asynchronous condition ( $p = .227$ ) or the unimodal auditory condition ( $p = .146$ ). Infants' reduced processing time in the unimodal auditory condition is likely attributable to the reduction in overall amount of stimulation.

We also examined whether infants' looking behaviors were differentially affected by prosodic passages specifying approval versus prohibition (see Table 4). We investigated this in two ways. First, we examined whether the direction of the prosody change (habituation to approval and test with prohibition or vice versa) affected visual recovery for infants who were able to discriminate a change in prosody (i.e., bimodal synchronous condition). We conducted a  $2 \times 2$  ANOVA with prosody change test

**Table 4**

Means (and standard deviations) for visual recovery (in seconds) by condition, prosody change test type, and test prosody.

	Prosody change			No prosody change		
	Approval	Prohibition	Overall	Approval	Prohibition	Overall
Bimodal synchronous	11.40 (8.23)**	2.71 (6.40)	7.055 (8.35)**	−5.03 (5.71)	0.06 (2.08)	−2.41 (4.93)
Unimodal auditory	3.91 (4.95)	2.81 (8.27)	3.36 (7.60)	1.95 (1.71)	3.15 (8.61)	2.55 (5.73)
Bimodal asynchronous	−0.15 (6.77)	−0.64 (5.93)	0.48 (7.16)	−6.22 (5.12)	4.62 (4.07)	0.40 (4.42)

Note. Mean visual recovery for each change group in each condition was compared with the no-change group in the corresponding condition. All significant comparisons also met the significance criteria for familywise error correction according to the multistage Bonferroni correction.

\*\*  $p < .01$ .

type (change or no change) and test prosody (approval or prohibition) as between-participants factors for visual recovery scores in the bimodal synchronous condition. In addition to the main effect of prosody change test type reported above, results indicated a significant interaction of prosody change test type and test prosody,  $F(1, 20) = 8.03$ ,  $p = .010$ ,  $\eta_p^2 = .29$ . Infants who received test trials specifying approval (but not prohibition) had significantly greater visual recovery to a change in prosody than to no change,  $t(10) = 4.02$ ,  $p = .002$ ,  $d = 2.32$  (significant when adjusted for multiple comparisons,  $p = .05/2 = .025$ ). This suggests that the main effect of prosody change test type was carried primarily by visual recovery to the approval prosody. However, the visual recovery scores for infants tested to approval versus prohibition did not differ for the unimodal auditory or bimodal asynchronous condition (see Table 4). Therefore, infants in these control conditions displayed no visual preference for approval prosody, demonstrating that the visual recovery indicating discrimination of prosody was specific to the synchronous audiovisual speech condition. It should also be noted, however, that sample size constrained our ability to detect visual recovery differences between prosody subgroups given that there were just 6 infants in each subgroup. Thus, no firm conclusions should be drawn from the presence or absence of a difference between the change and no-change conditions for infants tested to prosodic passages specifying prohibition given the low statistical power for this secondary analysis.

A second approach to investigating whether infants were differentially sensitive to prosody specifying approval versus prohibition was to assess overall processing time during habituation for each prosody for the full sample of infants. We conducted a one-way ANOVA on processing time with habituation prosody (approval or prohibition) as a between-participants factor. There was no significant difference in processing time for infants habituated to prosodies specifying approval versus prohibition ( $p = .79$ ). This did not differ as a function of condition ( $ps > .20$ ). Thus, there was no evidence that infants took longer to habituate to one prosody over the other.

## Discussion

The current study assessed whether intersensory redundancy could facilitate 4-month-old infants' ability to abstract prosodic information specifying approval versus prohibition in IDS. We predicted that intersensory redundancy provided by naturalistic, synchronous audiovisual speech would facilitate detection of prosody. According to the intersensory redundancy hypothesis, information that is presented redundantly and synchronously across sensory modalities facilitates detection of amodal properties. This is particularly true when a task is difficult relative to the abilities of the perceiver, as is the case early in development. Prosody is characterized by amodal properties—changes in temporal and intensity patterns common across auditory and visual speech. Typically, infants experience prosody in the context of IDS, and prosodic information can be detected both visually and acoustically. However, research has primarily investigated prosody as a vocal phenomenon, and so it is unknown how multimodal presentation of prosodic information affects infants' discrimination and categorization. Therefore, in the current study, infants were habituated to a woman reciting three phrases using a prosody characteristic of approval or prohibition followed by visual recovery test trials depicting the opposite prosody under conditions that provided intersensory redundancy (synchronous bimodal)

versus conditions that did not (unimodal visual and bimodal asynchronous). The current study yielded several important findings.

First, consistent with our main predictions, the findings demonstrate that intersensory redundancy facilitates infant detection of prosody and that at 4 months of age only infants who received naturalistic, synchronous audiovisual speech displayed detection of prosodic information. Infants in the bimodal synchronous condition who received a novel passage with a change in prosody exhibited greater visual recovery than infants who received a novel passage with no change in prosody. Infants in the nonredundant conditions (unimodal auditory and bimodal asynchronous) did not differ in their visual recovery between the prosody change and prosody no-change test types. Furthermore, in support of this hypothesis, findings indicated that only infants in the bimodal synchronous condition demonstrated visual recovery significantly above chance to a novel passage with a change in prosody. In contrast, infants who received a novel passage and a change in prosody in the bimodal asynchronous and unimodal auditory conditions showed no significant visual recovery to the prosody change.

The current study used a traditional ANOVA-based statistical approach with visual recovery as our dependent measure to assess infant discrimination (see Bahrick et al., 2002; Bahrick & Lickliter, 2000; Bahrick & Newell, 2008; Young-Browne, Rosenfeld, & Horowitz, 1977). An alternative approach that has also been used, and that has the advantage of taking into account infants' initial looking time, is a repeated-measures ANOVA with baseline, post-habituation, and test trial looking (see footnote 2). Using this approach, we also corroborated our main findings of infant discrimination of prosody (with an interaction of condition and trial type); infants showed longer looking to test trials than post-habituation trials in the synchronous speech condition but not in the unimodal auditory or asynchronous speech condition. This provides additional support for the conclusion that intersensory redundancy facilitates infant discrimination of prosody.<sup>3</sup>

Research by Spence and Moore (2003) demonstrated that 6-month-old infants, but not 4-month-old infants, could discriminate and categorize a change in prosody from approval to comfort or vice versa. Their study provided nonredundant auditory presentations of IDS. Here, we extend these findings to a different prosodic contrast, approval versus prohibition, and demonstrate that younger infants, at 4 months, are able to discriminate a change in prosody only in the presence of intersensory redundancy—audiovisual synchrony between the face and the voice, as in naturalistic speech. Thus, our findings indicate that detection of prosodic information conveying communicative intent is facilitated by the intersensory redundancy provided by face–voice synchrony in audiovisual speech. Furthermore, given findings by Spence and Moore (2003), it is likely that although younger (4-month-old) infants require intersensory redundancy to detect prosodic information, older (6-month-old) infants, who have more experience with speech, can do so without the support of intersensory redundancy, although this prediction needs to be tested directly.

Second, consistent with our predictions, our findings indicated that 4-month-old infants are able to generalize prosodic patterns across changes in speech passages in naturalistic, synchronous audiovisual speech. Infants in the bimodal synchronous condition exhibited greater visual recovery to a change in passage when it was accompanied by a change in prosody than when there was no change in prosody. Infants showed significant (relative to chance) visual recovery in response to a novel passage only when presented with *both* audiovisual face–voice synchrony and a change in prosody. In other words, they demonstrated invariant detection (Gibson, 1969) by abstracting an invariant prosodic pattern across changes in linguistic content in synchronous audiovisual speech. Thus, by 4 months of age, infants can detect invariant prosodic patterns (approval vs. prohibition) across changes in linguistic content (i.e., passage) only in the context of intersensory redundancy across the face and voice. Spence and Moore (2003) demonstrated that in unimodal auditory speech, older infants (at 6 months of age), but not younger infants (at 4 months of age), could categorize multiple exemplars of a given prosodic pattern and discriminate a novel exemplar only when there was a change in prosody. This was true for both naturalistic and low-pass filtered speech. Thus, 6-month-olds, but not 4-month-

<sup>3</sup> Multilevel modeling has also been used to analyze infant habituation data (Colombo & Mitchell, 2009; Liu & Spelke, 2017; Young & Hunter, 2015); however, this analytic strategy is best suited to addressing infant patterns of habituation and requires a much larger sample size than that of the current study.

olds, were able to detect invariant prosodic information across multiple tokens without the aid of intersensory redundancy. In contrast, the current study indicates that 4-month-olds were able to detect invariant prosodic information across changes in speech passages—only in the context of intersensory redundancy provided by synchronous audiovisual speech and not in its absence (in unimodal auditory or asynchronous audiovisual speech). Infants exhibited significant visual recovery to a change in passage when accompanied by a change in prosody in the synchronous audiovisual condition but not in the other conditions. Although methodologies and stimuli differed somewhat across the two studies (however, both used fluent naturalistic speech and tested prosody detection by assessing generalization to a novel speech token), taken together, these findings are suggestive of a developmental shift in the basis for detecting invariant prosodic information across multiple speech tokens. In early development infants rely on intersensory redundancy in audiovisual speech to detect invariant prosodic information, and later in development infants no longer need to rely on this information and can detect invariant prosodic patterns across multiple examples of unimodal auditory speech.

Third, findings from the current study indicated no evidence that infants were able to detect a change in passage alone under any condition. In all test trials, infants received a novel passage. Passages consisted of three short phrases with 14 or 15 syllables each. For half of the infants the novel passage was accompanied by a change in prosody, and for the other half it was not. Results indicated no evidence of visual recovery to a change in passage without a change in prosody in any condition (synchronous audiovisual, unimodal auditory, or asynchronous audiovisual). This is consistent with findings of previous research with infants aged 4–6 months (Moore et al., 1997; Spence & Moore, 2003).

Interestingly, our findings indicated little evidence that infants were differentially affected by prosodies of approval versus prohibition. Infants displayed no difference in their overall processing time across habituation as a function of whether they heard passages conveying approval versus prohibition, and this did not differ as a function of condition. There was, however, limited evidence of differential preference for approval versus prohibition in the group of infants who received synchronous faces and voices, but the small sample sizes in these subgroups ( $n = 6$ ) precludes drawing any firm conclusions. Infants in the subgroup who received test trials specifying approval showed greater visual recovery to a change in prosody than to no change, indicating detection of the novel prosody when it was approval. In contrast, this difference was not evident in the subgroup of infants who received test trials specifying prohibition. Although this indicates that the main effects of visual recovery to a prosody change were carried by infants who received test trials specifying approval, there were too few participants in each subgroup to draw any conclusions about preferences for one prosody over the other. Note also that differential preferences are often obtained when one stimulus is more salient or positive than another (greater visual recovery to negative/sad affect followed by tests with positive/happy affect than to positive/happy affect followed by tests with negative/sad affect; see Walker-Andrews & Lennon, 1991, for examples). Positive affect is typically more reinforcing and attractive. These additional factors may contribute to the visual recovery patterns in the current study.

Finally, there were differences in processing time between the bimodal synchronous and asynchronous conditions. Infants in the bimodal asynchronous condition exhibited marginally greater looking during baseline than infants in the bimodal synchronous condition. However, this increased opportunity to process prosodic information in the asynchronous audiovisual condition—in the absence of intersensory redundancy provided by synchronous faces and voices—did not translate to detecting a change in prosody. The asynchronous condition provides a control for the synchronous condition by equating the overall amount and type of stimulation and varying only the temporal synchrony between them. That infants detected a change in prosody in synchronous audiovisual speech, but not in asynchronous audiovisual speech, indicates the unique role of audiovisual temporal synchrony in facilitating attention and detection of prosody. Furthermore, these findings demonstrated infant detection of prosodic changes on the basis of significantly less overall processing time, highlighting the efficiency of intersensory redundancy in promoting attention to amodal properties.

In the natural environment, infants typically experience the prosodic patterns present in IDS in the context of multimodal speech with face–voice synchrony. However, prior research on the early development of prosody detection has focused on speech as an auditory stream (Cooper & Aslin, 1990;

Moore et al., 1997; Soderstrom, 2007; Spence & Moore, 2003; Trainor et al., 2000). The current study demonstrates the importance of investigating the detection of prosodic information using multimodal audiovisual speech. Audiovisual speech provides a host of temporal and intensity pattern information (amodal information) invariant across visual and auditory speech that facilitates detection of prosodic information.

Prior research has demonstrated the salience of IDS to infants and has highlighted several of its functions in facilitating social and language development (Colombo et al., 1995; Cooper & Aslin, 1990; Fernald, 1985). Little research, however, has focused on discrimination of prosodic differences in IDS and on the conditions that facilitate detection early in development. The current study demonstrates that the intersensory redundancy present in naturalistic audiovisual speech aids young infants in discriminating communicative intent in spoken language. Intersensory redundancy appears to bootstrap infants' ability to perceive and distinguish prosodic information, serving as one of the first bases for perceiving meaning in fluent speech.

## Acknowledgments

This work was supported by National Institutes of Health (NIH) Grants K02-HD064943 and R01-HD053776 awarded to the first author and by an American Psychological Association (APA) PRIME fellowship and NIH/NIGMS (National Institute of General Medical Sciences) Grant R25 GM061347 awarded to the fourth author. We thank Mariana Vaillant-Molina, Ana Bravo, Melissa Argumosa, and Laura Batista for assistance in data collection.

## References

- Bahrick, L. E., Flom, R., & Lickliter, R. (2002). Intersensory redundancy facilitates discrimination of tempo in 3-month-old infants. *Developmental Psychobiology*, 41, 352–363.
- Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology*, 36, 190–201.
- Bahrick, L. E., & Lickliter, R. (2002). Intersensory redundancy guides early perceptual and cognitive development. In R. Kail (Ed.), *Advances in child development and behavior* (Vol. 30, pp. 153–187). San Diego: Academic Press.
- Bahrick, L. E., & Lickliter, R. (2014). Learning to attend selectively: The dual role of intersensory redundancy. *Current Directions in Psychological Science*, 23, 414–420.
- Bahrick, L. E., & Newell, L. C. (2008). Infant discrimination of faces in naturalistic events: Actions are more salient than faces. *Developmental Psychology*, 44, 983–996.
- Bryant, G. A., & Barrett, H. C. (2007). Recognizing intentions in infant-directed speech: Evidence for universals. *Psychological Science*, 18, 746–751.
- Caron, A. J., Caron, R. F., & MacLean, D. J. (1988). Infant discrimination of naturalistic emotional expressions: The role of face and voice. *Child Development*, 59, 604–616.
- Colombo, J., Frick, J. E., Ryther, J. S., Coldren, J. T., & Mitchell, D. W. (1995). Infants' detection of analogs of "motherese" in noise. *Merrill-Palmer Quarterly*, 41, 104–113.
- Colombo, J., & Mitchell, D. W. (2009). Infant visual habituation. *Neurobiology of Learning and Memory*, 92, 225–234.
- Cooper, R. P., Abraham, J., Berman, S., & Staska, M. (1997). The development of infants' preference for motherese. *Infant Behavior and Development*, 20, 477–488.
- Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, 61, 1584–1595.
- Fernald, A. (1984). The perceptual and affective salience of mothers' speech to infants. In L. Feagans, C. Garvey, & R. Golinkoff (Eds.), *The origins and growth of communication* (pp. 5–29). Norwood, NJ: Ablex.
- Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, 8, 181–195.
- Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, 60, 1497–1510.
- 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.
- Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, 16, 477–501.
- Flom, R., & Bahrick, L. E. (2007). The development of infant discrimination of affect in multimodal and unimodal stimulation: The role of intersensory redundancy. *Developmental Psychology*, 43, 238–252.
- Gibson, E. J. (1969). *Principles of perceptual learning and development*. East Norwalk, CT: Appleton-Century-Crofts.
- Golinkoff, R. M., & Alioto, A. (1995). Infant-directed speech facilitates lexical learning in adults hearing Chinese: Implications for language acquisition. *Journal of Child Language*, 22, 703–726.



- Grieser, D. L., & Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology*, 24, 14–20.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65–70.
- Horowitz, F. D., Paden, L., Bhana, K., & Self, P. (1972). An infant-control procedure for studying infant visual fixation. *Developmental Psychology*, 7, 90.
- Horowitz, F. D. (Ed.). (rowitz 1975). *Visual attention, auditory stimulation, and language discrimination in young infants. Monographs of the Society for Research in Child Development* (39(5–6), pp. 1–140). <https://doi.org/10.2307/1165968>.
- Jaccard, J. J., & Guilamo-Ramos, V. (2002). Analysis of variance frameworks in clinical child and adolescent psychology: Issues and recommendations. *Journal of Clinical Child and Adolescent Psychology*, 31, 130–146.
- Johnstone, T., & Scherer, K. R. (2000). Vocal communication of emotion. In M. Lewis & J. Haviland (Eds.), *The handbook of emotion* (pp. 220–235). New York: Guilford.
- Juslin, P. N., & Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion*, 1, 381–412.
- Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, 129, 770–814.
- Kaplan, P. S., Goldstein, M. H., Huckleby, E. R., & Cooper, R. P. (1995). Habituation, sensitization, and infants' responses to motherese speech. *Developmental Psychobiology*, 28, 45–57.
- Ladd, R. D., Silverman, K. E. A., Tolkmitt, F., Bergmann, G., & Scherer, K. R. (1985). Evidence for the independent function of intonation contour type, voice quality, and F0 range in signaling speaker affect. *Journal of the Acoustical Society of America*, 78, 435–444.
- Lewkowicz, D. J. (1996). Perception of auditory–visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 1094–1106.
- Liu, S., & Spelke, E. S. (2017). Six-month-old infants expect agents to minimize the cost of their actions. *Cognition*, 160, 35–42.
- Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word learning in infant- and adult-directed speech. *Language Learning and Development*, 7, 185–201.
- Moore, D. S., Spence, M. J., & Katz, G. S. (1997). Six-month-olds' categorization of natural infant-directed utterances. *Developmental Psychology*, 33, 980–989.
- Morgan, J. L. (1996). Prosody and the roots of parsing. *Language and Cognitive Processes*, 11, 69–106.
- Nazzi, T., Kemler Nelson, D. G., Jusczyk, P. W., & Jusczyk, A. M. (2000). Six-month-olds' detection of clauses embedded in continuous speech: Effects of prosodic well-formedness. *Infancy*, 1, 123–147.
- Newport, E., Gleitman, H., & Gleitman, L. (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children* (pp. 109–149). Cambridge, UK: Cambridge University Press.
- Papoušek, M., Bornstein, M. H., Nuzzo, C., Papoušek, H., & Symmes, D. (1990). Infant responses to prototypical melodic contours in parental speech. *Infant Behavior and Development*, 13, 539–545.
- Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2014). Look who's talking: Speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science*, 17, 880–891.
- Saint-Georges, C., Chetouani, M., Cassel, R., Apicella, F., Mahdhaoui, A., Muratori, F., ... Cohen, D. (2013). Motherese in interaction: At the cross-road of emotion and cognition? (A systematic review). *PLoS One*, 8(10) e78103.
- Sakalou, E., & Gattis, M. (2012). Infants infer intentions from prosody. *Cognitive Development*, 27, 1–16.
- Santarcangelo, S., & Dyer, K. (1988). Prosodic aspects of motherese: Effects on gaze and responsiveness in developmentally disabled children. *Journal of Experimental Child Psychology*, 46, 406–418.
- Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, 99, 143–165.
- Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40, 227–256.
- Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology*, 18, 668–671.
- Singh, L., Morgan, J. L., & Best, C. T. (2002). Infants' listening preferences: Baby talk or happy talk? *Infancy*, 3, 365–394.
- Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, 27, 501–532.
- Song, J. Y., Demuth, K., & Morgan, J. (2010). Effects of the acoustic properties of infant-directed speech on infant word recognition. *Journal of the Acoustical Society of America*, 128, 389–400.
- Spence, M. J., & Moore, D. S. (2003). Categorization of infant-directed speech: Development from 4 to 6 months. *Developmental Psychobiology*, 42, 97–109.
- Spinelli, M., Fasolo, M., & Mesman, J. (2017). Does prosody make the difference? A meta-analysis on relations between prosodic aspects of infant-directed speech and infant outcomes. *Developmental Review*, 44, 1–18.
- Trainor, L. J., Austin, C. M., & Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychological Science*, 11, 188–195.
- Walker-Andrews, A. S., & Grolnick, W. (1983). Discrimination of vocal expressions by young infants. *Infant Behavior and Development*, 6, 491–498.
- Walker-Andrews, A. S., & Lennon, E. (1991). Infants' discrimination of vocal expressions: Contributions of auditory and visual information. *Infant Behavior and Development*, 14, 131–142.
- Werker, J. F., & McLeod, P. J. (1989). Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, 43, 230–246.
- Young, D. S., & Hunter, D. R. (2015). Random effects regression mixtures for analyzing infant habituation. *Journal of Applied Statistics*, 42, 1421–1441.
- Young-Browne, G., Rosenfeld, H. M., & Horowitz, F. D. (1977). Infant discrimination of facial expressions. *Child Development*, 48, 555–562.