Developmental Science
Developmental Science 19:2 (2016), pp 318–328 DOI: 10.1111/desc.12298

PAPER

When infants talk, infants listen: pre-babbling infants prefer listening to speech with infant vocal properties

Matthew Masapollo,[1,2] Linda Polka[1,2] and Lucie Ménard[2,3]

1. *School of Communication Sciences & Disorders, McGill University, Canada*
2. *Centre for Research on Brain, Language & Music, McGill University, Canada*
3. *Département de Linguistique, Université du Québec à Montréal, Canada*

Abstract

To learn to produce speech, infants must effectively monitor and assess their own speech output. Yet very little is known about how infants perceive speech produced by an infant, which has higher voice pitch and formant frequencies compared to adult or child speech. Here, we tested whether pre-babbling infants (at 4–6 months) prefer listening to vowel sounds with infant vocal properties over vowel sounds with adult vocal properties. A listening preference favoring infant vowels may derive from their higher voice pitch, which has been shown to attract infant attention in infant-directed speech (IDS). In addition, infants' nascent articulatory abilities may induce a bias favoring infant speech given that 4- to 6-month-olds are beginning to produce vowel sounds. We created infant and adult /i/ ('ee') vowels using a production-based synthesizer that simulates the act of speaking in talkers at different ages and then tested infants across four experiments using a sequential preferential listening task. The findings provide the first evidence that infants preferentially attend to vowel sounds with infant voice pitch and/or formants over vowel sounds with no infant-like vocal properties, supporting the view that infants' production abilities influence how they process infant speech. The findings with respect to voice pitch also reveal parallels between IDS and infant speech, raising new questions about the role of this speech register in infant development. Research exploring the underpinnings and impact of this perceptual bias can expand our understanding of infant language development.

Research highlights

- How infants perceive infant speech signals has been grossly neglected, despite its significance for speech and language development.
- This is the first documented evidence that infants prefer listening to infant vowel sounds over adult vowel sounds.
- The perceptual bias uncovered here may arise from infants' interest in their own early vocalizations, and may be shaping how adults speak to infants.

Introduction

For infants to learn to produce speech, they must effectively monitor and assess their own self-generated speech (Kuhl & Meltzoff, 1996; Rvachew, Slawinski, Williams & Green, 1996; Rvachew, Mattock, Polka & Ménard, 2006; Doupe & Kuhl, 2008; Polka, Masapollo & Ménard, 2014). Despite the vital role of this skill in forging early links between the perceptual and motor systems for speech, very little is known about how infants perceive infant-produced speech. Research investigating the development of infant speech production, on the other hand, shows that by 10 months, infants exposed to different languages produce babbling with some language-specific vowel characteristics when measured acoustically (de Boysson-Bardies, Halle, Sagart & Durand, 1989; Rvachew *et al*., 1996, 2006; Alhaidary & Rvachew, 2011). In addition, 5-month-old infants have been shown to rapidly modify their vocalizations in response to audio-visual recordings of vowels produced (non-interactively) by an adult on a television (Kuhl &

Address for correspondence: Linda Polka, School of Communication Sciences and Disorders, 2001 McGill College, 8th floor McGill University, Montreal, Quebec, Canada H3A 1G1; e-mail: linda.polka@mcgill.ca

Meltzoff, 1996). These findings suggest that young infants have some ability to process their *own* vocal output, and can link sensory patterns that they have seen and heard with sensory-motor patterns that they are attempting to produce.

One factor that may facilitate infants' rapid speech production learning is a perceptual bias favoring infant speech. In the present study, we use advances in speech synthesis to investigate whether *pre-babbling* infants display a perceptual bias for listening to infant vowel sounds over adult vowel sounds. Several factors might lead one to predict that an infant vowel bias will be observed.

First, infant speech has acoustic properties that are known to be perceptually salient to infants. Because an infant's vocal folds are much shorter and lighter than those of an adult or even a toddler, the infant voice pitch range is much higher. Fundamental frequency (f0) provides an acoustic measure of voice pitch, reflecting the rate at which the vocal folds open and close during vibration. As shown in Figure 1, average f0 values (during vowel production) decrease with age, with a sharp decline between birth and 2 years of age coinciding with this period of very rapid growth (Vorperian & Kent, 2007; see also Table 1).

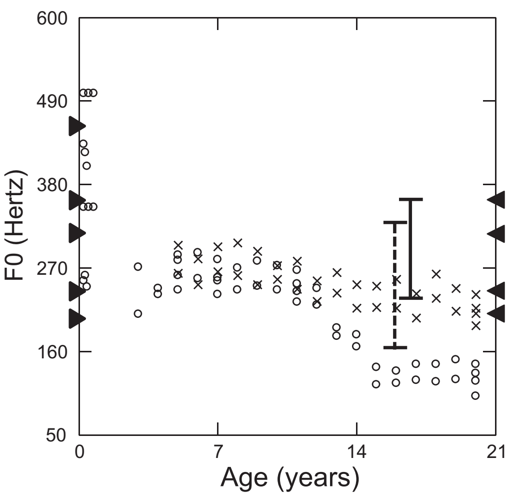

Figure 1 *The typical mean voice pitch (f0) values (during vowel production) for speakers across the life span. Values are plotted with a circle (o) for male speakers, and with a cross (x) for female speakers producing adult-directed speech (ADS). The values implemented in the present study are indicated as triangles; vowels with infant formants (pointing right) and vowels with adult formants (pointing left). The typical pitch ranges observed for female speakers for ADS (dotted line) and for IDS (solid line) are shown for comparison. These data are drawn from multiple studies (see Table 1).*

During speech production, the acoustic energy created by vocal fold vibration is modified as it resonates in the vocal tract, the tube formed by the vocal folds on one end and the lips on the other. The vocal tract resonances concentrate acoustic energy into discrete frequency bands called formants, which are numbered F1, F2, F3, etc., going from low to high frequencies. Vocal resonances (and their associated formants) are inversely related to vocal tract length, i.e. a shorter vocal tract gives rise to higher formant frequencies. The first (F1) and second (F2) formant frequencies provide critical information for the perception of vowel quality (Strange, 1999). Traditionally, F1 and F2 are plotted as shown in Figure 2 with the vowels /i/ ('ee'), /a/ ('ah'), and /u/ ('oo') forming an acoustic space resembling a triangle. This vowel triangle encompasses all possible vowel sounds for a given vocal tract length. Because an infant's vocal tract is much shorter compared to an adult or child, their vowel formants are also much higher in frequency (Vorperian & Kent, 2007). Figure 2 shows typical acoustic vowel spaces observed for an adult, child, and infant speaker. It is clear that the infant vowel space has notably higher F1 and F2 values.

In many languages caregivers increase their voice pitch and expand their vowel space when producing infant-directed speech (IDS; Fernald, Taeschner, Dunn, Papousek, de Boysson-Bardies *et al.*, 1989; Kuhl, Andruski, Chistovich, Chistovich, Kozhevnikova *et al.*, 1997). In addition, these acoustic modifications have been shown to attract infant attention (Fernald & Kuhl, 1987; Zhang, Koerner, Miller, Grice-Patil, Svec *et al.*, 2011). The f0 range typically found in IDS (for female adults), indicated in Figure 1, overlaps with infant speech, showing clear parallels between the f0 properties of infant speech and IDS. In contrast, the vowel space corresponding to IDS speech (female adult), plotted in Figure 2, fails to reveal an overlap between IDS and infant speech with respect to formant patterns. In the infant vowel space, all formants are shifted to higher frequencies, whereas in IDS, the vowel space is stretched to exaggerate the corner vowels; this shifts the formants in different directions, e.g. when /i/ is produced in IDS, F1 is lower and F2 is higher compared to a non-IDS production. Thus, on the basis of infants' response to IDS we predict that infants will prefer speech with infant voice pitch values; however, this literature provides no basis for predicting how infants will respond to infant vowel formants.

A second factor that could influence infants' perception of infant speech is their *own* emerging vocal production patterns, especially those phonetic dimensions or forms that they are actively learning to control. Support for this position comes from recent studies by

Table 1 *Summary table of studies reporting average voice pitch (f0) values (during vowel production) across the lifespan*

Speaker Age, Sex	Sample Size	Mean f0 (Hz)	f0 Range (Hz)	Study	Corpus
Infants 3 ms	n=7	445	350-500	Kent & Murray (1982)	f0 values taken from spontaneous infant vowel-like vocalizations
Infants 3 ms	n=24	325	253-433	Kuhl & Meltzoff (1996)	f0 values taken from infant vocal imitations of adult (female) /i/, /a/, and /u/ recordings
Infants 4 ms	n=24	323	260-422	Kuhl & Meltzoff (1996)	f0 values taken from infant vocal imitations of adult (female) /i/, /a/, and /u/ recordings
Infants 5 ms	n=24	313	245-404	Kuhl & Meltzoff (1996)	f0 values taken from infant vocal imitations of adult (female) /i/, /a/, and /u/ recordings
Infants 6 ms	n=7	450	350-500	Kent & Murray (1982)	f0 values taken from spontaneous infant vowel-like vocalizations
Infants 9 ms	n=7	415	350-500	Kent & Murray (1982)	f0 values taken from spontaneous infant vowel-like vocalizations
Children 3 yrs	n=10	232	209-271	Assman & Katz (2000)	f0 values taken from 12 English vowels; male and female data collapsed for child speakers.
Children 4 yrs	n=20	240	235-243	Perry, Ohde, & Ashmead (2001)	f0 values taken from 7 English vowels; male and female data collapsed for child speakers.
Children 5 yrs	n=10	259	241-280	Assman & Katz (2000)	f0 values taken from 12 English vowels; male and female data collapsed for child speakers.
Children 5 yrs male	n=19	267	260-287	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Children 5 yrs female	n=13	274	263-300	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Children 6 yrs male	n=11	319	256-290	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Children 6 yrs female	n=16	263	248-282	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Children 7 yrs	n=10	248	235-257	Assman & Katz (2000)	f0 values taken from 12 English vowels; male and female data collapsed for child speakers.
Children 7 yrs male	n=11	265	253-281	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Children 7 yrs female	n=24	277	265-298	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Children 8 yrs male	n=25	251	241-270	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Children 8 yrs female	n=25	274	260-303	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Children 9 yrs male	n=23	259	246-279	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Children 9 yrs female	n=23	264	248-292	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Children 10 yrs male	n=25	253	241-273	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Children 10 yrs female	n=14	262	255-273	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Children 10-12 years	n=46	236	225-249	Hillenbrand, Getty, Clark & Wheeler (1995)	f0 values taken from 12 English vowels; male and female data collapsed for child speakers.
Children 11 yrs male	n=24	249	239-267	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Children 11 yrs female	n=24	250	242-279	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Children 12 yrs male	n=22	229	221-243	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Children 12 yrs female	n=22	234	226-253	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Children 13 yrs male	n=16	185	180-191	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Children 13 yrs female	n=16	247	237-264	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Children 14 male	n=11	173	166-182	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Children 14 yrs female	n=11	226	217-248	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Children 15 yrs male	n=11	127	116-139	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Children 15 yrs female	n=11	226	218-246	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Children 16 yrs male	n=11	125	118-134	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Children 16 yrs female	n=11	231	217-255	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Children 17 yrs male	n=10	129	122-143	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Children 17 yrs female	n=10	214	204-236	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Children 18 yrs male	n=10	127	120-143	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Children 18 yrs female	n=10	242	230-262	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Adult male (ADS)	n=29	133	123-149	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Adult female (ADS)	n=27	227	213-243	Lee, Potamianos, & Narayanan (1999)	f0 values taken from 10 English vowels
Adult male (ADS)	n=45	120	121-143	Hillenbrand, Getty, Clark & Wheeler (1995)	f0 values taken from 12 English vowels
Adult female (ADS)	n=48	220	210-235	Hillenbrand, Getty, Clark & Wheeler (1995)	f0 values taken from 12 English vowels
Adult male (ADS)	n=10	135	101-131	Assman & Katz (2000)	f0 values taken from 12 English vowels
Adult female (ADS)	n=10	205	194-217	Assman & Katz (2000)	f0 values taken from 12 English vowels
Adult female (ADS)	n=48	234	159-335	Lam & Kitamura (2012)	f0 values taken from /i/, /a/, and /u/
Adult female (IDS)	n=48	309	247-362	Lam & Kitamura (2012)	f0 values taken from /i/, /a/, and /u/
Adult female (ADS)	n=32	219	NA	Liu, Kuhl & Tsao (2003)	f0 values taken from /i/, /a/, and /u/
Adult female (IDS)	n=32	274	NA	Liu, Kuhl & Tsao (2003)	f0 values taken from /i/, /a/, and /u/

Table 1 *(Continued)*

Speaker Age, Sex	Sample Size	Mean f0 (Hz)	f0 Range (Hz)	Study	Corpus
Adult female (IDS)	n=22	232	NA	Danielson, Siedhl, Onishi, Alamian & Cristia (2014)	f0 values taken from /i/, /a/, and /u/ - Monolingual English mothers
Adult female (IDS)	n=20	289	NA	Danielson, Siedhl, Onishi, Alamian & Cristia (2014)	f0 values taken from /i/, /a/, and /u/ - Monolingual French mothers
Adult female (IDS)	n=9	236	NA	Danielson, Siedhl, Onishi, Alamian & Cristia (2014)	f0 values taken from /i/, /a/, and /u/ - Bilingual English-French mothers in English mode
Adult female (IDS)	n=9	260	NA	Danielson, Siedhl, Onishi, Alamian & Cristia (2014)	f0 values taken from /i/, /a/, and /u/ - Bilingual English-French mothers in French mode

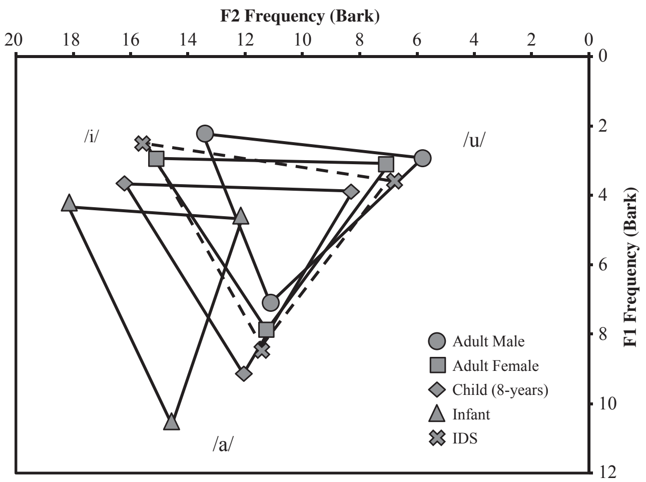

Figure 2 *Example vowel spaces corresponding to an adult male, adult female, child (at 8 years), and infant speaker; the corner vowels /i/ ('ee'), /a/ ('ah'), and /u/ ('oo') are labeled. The exaggerated vowel space of infant-directed speech (IDS; as spoken by a female) is also indicated for comparison. The first formant frequency (F1) is plotted on the y-axis and the second formant frequency (F2) on the x-axis (note that formant frequencies values increase going from right to left for F2 and going from top to bottom for F1). The frequency axes are scaled in Bark units, which transform formant frequencies into units of equal psychophysical distance. Note: to convert from Bark to Hz: F(Hz) = 650*sinh(F(bark)/7); from Hz to Bark: F (bark) = 7*asinh(F(Hz) /650). See also Schroeder, Atal and Hall (1979).*

Vihman and colleagues, which show that infants who are producing canonical babbling (at 9–11 months) are more attentive to isolated non-words and sentences containing consonants that they are beginning to produce in a stable form (DePaolis, Vihman & Keren-Portnoy, 2011; DePaolis, Vihman & Nakai, 2013; Majorano, Vihman & DePaolis, 2013). These findings led Vihman *et al.* to conclude that an infant's speech *output* biases their speech *intake*; that is, an infant's familiarity with and interest in their own emerging production patterns make segments in the input speech that resemble those patterns more perceptually salient. Infants begin producing vowels and vowel-like sounds before they are able to produce the well-formed syllables that characterize canonical babbling (Kent, 1992; Rvachew *et al.*, 1996, 2006). If infant 'output biases intake' as soon as infants begin to formulate speech sounds, then one would expect to find a perceptual bias favoring infant vowels once vowel production is under way in the pre-babbling stage (at 4–6 months). From this view, one would predict that pre-babbling infants would prefer *both* the voice pitch and the formants of infant vowels since these properties jointly form infant speech output.

Here, we present a series of experiments designed to directly test whether 4- to 6-month-olds preferentially attend to vowels with infant vocal properties over vowels with adult vocal properties. Using synthesized infant and adult /i/ ('ee') vowels, we measured infant responsiveness to the infant and adult vowels using a sequential preferential listening task (Polka, Jusczyk & Rvachew, 1995). We selected /i/ because this vowel occurs in all languages (Lindblom, 1986) and is uniformly perceived as the same vowel across many languages including English and French, the majority languages of Montreal. We presented isolated vowel sounds, which are not unusual or unrealistic speech signals for 4- to 6-month-olds. At this age infants begin to produce isolated vowels and vowel-like sounds, and caregivers will often imitate their infants and produce isolated vowels along with more complex utterances (Pawlby, 1977; Masakata, 2003).

Experiment 1

Methods

Participants

Data were analyzed for 18 infants, aged 4 to 6 months (7 male, M = 159 days, R = 132–177 days). In this and in all subsequent experiments all of the infants were from families who speak English or French and were full-term with no known health problems. Four additional infants were excluded due to fussiness (2), experimental error (1) and mean looking time (for each vowel type) that was more than 2 standard deviations above or below the group mean (1). The latter exclusion was made (in all experiments) so that group averages would not be skewed by extreme listening times of an individual infant.

Stimuli

The speech stimuli consisted of four isolated /i/ vowels synthesized using the Variable Linear Articulatory Model (VLAM), described by Ménard, Schwartz and Boe (2004). VLAM simulates speakers across a broad age range from infancy to adulthood. The infant vowels had typical pitch and formants values for an infant; adult vowels had typical pitch and formants values for a female adult. Two /i/ vowels differing only in pitch were created for each talker type: 210 and 240 Hz for the female tokens, and 360 and 450 Hz for the infant tokens. Details of the VLAM synthesis and an acoustic description of the vowels are provided in the supporting information (SI). Spectrograms showing examples of the vowel stimuli used in each experiment are shown in Figure 3.

All vowels were 500 ms long and matched in intensity and intonation contour. The stimuli were judged to be highly intelligible, natural-sounding exemplars of /i/ ('ee') spoken by adult female and infant talkers by English-speaking and French-speaking adults. For testing, we created four stimulus files (one per vowel) using

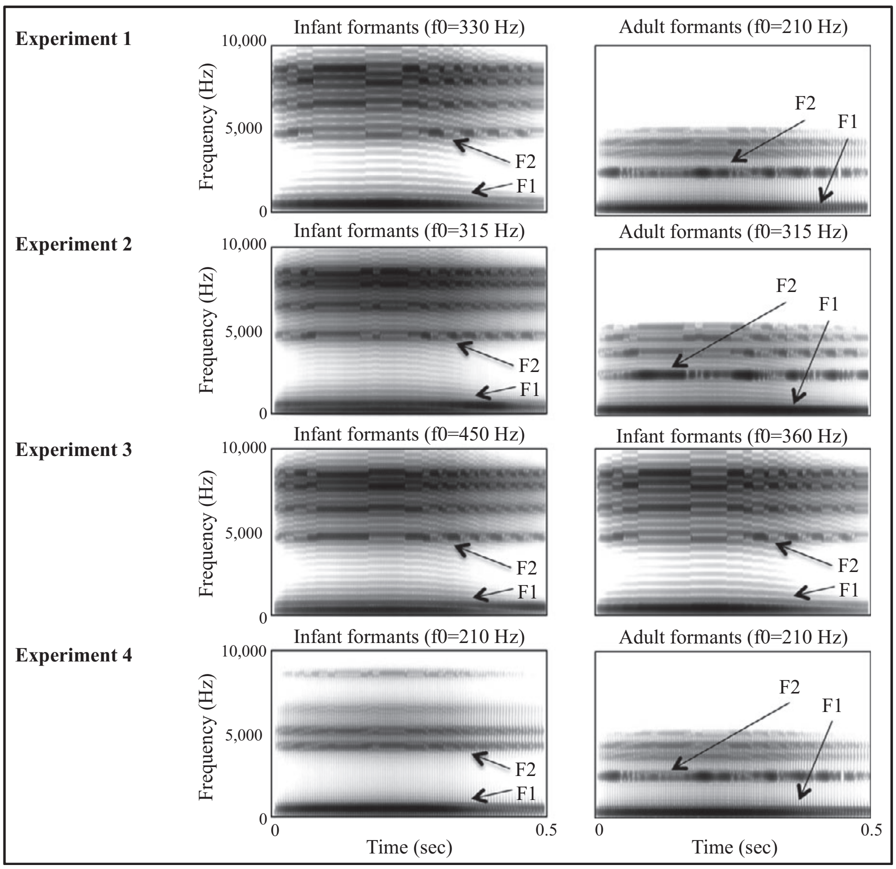

Figure 3 *Example spectrograms of /i/ vowels used in Experiments 1–4.*
Experiment 1: Infant vs. adult formants and voice pitch.
Experiment 2: Infant vs. adult formants; matching voice pitch values using f0 values appropriate for infant or IDS (female adult) speech.
Experiment 3: High vs. low voice pitch values both within infant range; infant formants.
Experiment 4: Infant vs. adult formants; matching voice pitch values using f0 values appropriate for adult female speech.

Praat (Boersma & Weenink, 2013); each file was 30 seconds including 20 repetitions of the same vowel with a 1000 ms inter-stimulus interval (ISI).

Procedure

Infants were tested using the look-to-listen procedure in which infants gain access to audio stimuli by fixating on a static visual pattern. Infants sat on their caregiver's lap at a distance of about 150 cm facing a 21-inch television monitor in a dimly lit curtained soundproof booth. Audio TRAK BSI-90 loudspeakers and a Sony digital video camera were located behind the curtain just below the TV screen. An experimenter observed the infant (outside the room) on a video monitor. The caregiver wore noise-canceling headphones and listened to music throughout the procedure to avoid influencing the infant's behavior. Experimental stimuli were presented and looking/listening times monitored using the software Habit 2000 (Cohen, Atkinson & Chaput, 2000). To start each trial, a red flashing light was presented to direct infants' attention followed by a black-and-white-checkerboard. The experimenter (who could not hear the stimuli) pressed a key when the infant fixated on the checkerboard, which activated the presentation of the auditory stimulus, providing an index of the infant's listening time. When the infant looked away for more than 2 seconds, the sound stopped and the screen went black. The minimum look time for a trial was 1 second. If the infant looked away for less than 2 seconds, then the sound continued to play, but the look away time was not included in the looking time for that trial. The trial terminated when the infant looked away for more than 2 seconds or the complete stimulus file (30 s long) had played. After a brief pause, the attention-getter returned to start the next trial.

Design

The experiment consisted of three consecutive phases: pre-test, preference test, and post-test. The pre- and post-test consisted of one trial of an instrumental music file; these trials were not analyzed. The pre-test was a warm-up to demonstrate the task to infants and the post-test confirmed that infants were attentive at the end of the test session. During the preference test phase, infants were presented with six trials of the vowel /i/ produced by the infant (three trials with the 360 Hz pitch token and three with the 450 Hz pitch token) and six trials of the vowel produced by the adult female (three trials with the 210 Hz pitch token and three with the 240 Hz pitch token). Vowel type was alternated across test trials; the vowel type presented first (adult or infant) was counterbalanced across participants.

Results

As is standard with this task, infants' mean listening time (MLT) to each vowel type (infant; adult) was calculated after removing the first preference test trial (Cooper & Aslin, 1994; Vouloumanos & Werker, 2004).[1] MLT scores were analyzed using a mixed two-way analysis of variance (ANOVA) with vowel type (infant vs. adult) as a within-subjects factor, and test trial order (infant first vs. adult first) as a between-subjects factor (see Figure 4). There was a main effect of vowel type, $F(1, 16) = 16.363$, $p = .001$, $\eta_p^2 = .506$; infants listened longer to infant ($M = 10.9$, $SD = 3.5$) than to adult vowels ($M = 8.2$, $SD = 2.2$). There was no main effect of test order, $F(1, 16) = .269$, $p = .611$, or two-way interaction, $F(1, 16) = .953$, $p = .343$.

Discussion

The findings in Experiment 1 confirm that infants display a perceptual bias favoring infant vowels over adult vowels. However, it is not clear which property of infant vowels contributes to this preference because the infant and adult vowels differed in both voice pitch and formants. Across the next three experiments we manipulate these dimensions to identify the properties contributing to this preference.

Experiment 2

To determine whether voice pitch properties support the listening preference observed in Experiment 1, we tested infants' preference for /i/ vowels synthesized with infant versus adult formant frequencies and *matching* voice pitch values. The f0 values selected, 315 Hz and 360 Hz, are characteristic of infant speech, but also fall within the f0 range reported for female adults (English) producing IDS (see Figure 1). If voice pitch information alone supports infants' listening preferences in Experiment 1, then no preference should be observed when voice pitch does not differ.

Methods

Participants

Data were analyzed from 22 infants, aged 4 to 6 months (14 male, M = 169 days, R = 139–227 days). Four additional infants were tested but excluded due to

[1] In each experiment, analyses were repeated with all test trials included; the same pattern of results was obtained in every case.

fussiness (1), experimenter error (1) and MLTs that were outliers with respect to the group mean (2).

Stimuli

The stimuli were four isolated /i/ vowels synthesized using the same parameters as the infant and female vowels in Experiment 1, except that the two infant tokens and the two female tokens had the same pitch values – 315 Hz and 360 Hz. Otherwise, the stimulus files were the same as in Experiment 1.

Procedure

The procedure was the same as in Experiment 1.

Design

The design was the same as in Experiment 1, except that test trials included six trials where the /i/ vowel had infant formant frequencies (three with f0 of 315 Hz; three with f0 of 360 Hz) and six trials where vowel /i/ vowel had adult formant frequencies (three with f0 of 315 Hz; three with f0 of 360 Hz). Vowel type was alternated across test trials; the vowel type presented first (infant or adult) was counterbalanced across participants and the voice pitch value presented first (360 or 315 Hz) was also counterbalanced within each test order.

Results

As in Experiment 1, MLT for each vowel type (adult vs. infant) was calculated without the first test trial, and the MLT scores were analyzed in a vowel type (infant vs. adult female) by test trial order ANOVA (see Figure 4). There was no main effect of vowel type, $F(1, 20) = .062$, $p = .806$, test order, $F(1, 20) = 2.446$, $p = .134$, or two-way interaction, $F(1, 20) = 1.049$, $p = .318$.

Discussion

In Experiment 2, infants did *not* listen longer to vowels with infant formant frequencies versus adult formant frequencies when the vowels had matching voice pitch values that are characteristic of an infant and fall within the range of IDS. This shows that the high voice pitch contributes to the listening preference observed in Experiment 1. However, it is not clear from the findings of Experiments 1 and 2 whether infants have a bias favoring infant-like pitch values or a general bias favoring high frequency signals. Although recent research on infant perception of musical tones points to a general bias (Marie & Trainor, 2012), research with speech fails to uphold this view. For example, a general auditory bias would predict that infants would prefer female to male speech, but the literature does not support this (Fernald & Kuhl, 1987). In addition, recent studies show that infants prefer listening to human speech over other natural complex sounds that are higher in pitch (Polka, Rvachew & Molnar, 2008; Vouloumanos, Hauser, Werker & Martin, 2010; Shultz & Vouloumanos, 2010). In Experiment 3 we address this issue in the context of vowel perception.

Experiment 3

In Experiment 3, we tested whether infants' preference for the infant vowels in Experiment 1 was driven by voice pitch *per se*, when voice pitch is varied within a range that is characteristic of an infant speaker. To do so, we tested infant preference for vowels with high versus low voice pitch values, when the vowels are matched to have the same infant formant frequencies. We expect infants to prefer the vowels with higher voice pitch if they have a general bias favoring higher f0 values. However, if infants fail to show a preference here, then this would suggest that infants' bias for the infant vowels cannot be explained as a simple preference for higher voice pitch.

Methods

Participants

Data were analyzed from 18 infants, aged 4 to 6 months (9 male, $M = 171$ days, $R = 139–182$ days). Four additional infants were tested but excluded due to fussiness (2) and MLTs that were outliers with respect to the group mean (2).

Stimuli

The stimuli were the same infant /i/ tokens used in Experiment 1 with f0 values of 360 Hz and 450 Hz.

Procedure

The procedure was the same as in Experiment 1.

Design

The design from Experiment 1 was used, except that test trials included six trials where the /i/ vowel had infant formant frequencies with a pitch of 450 Hz and six trials where the /i/ vowel had infant formant frequencies with a pitch of 360 Hz. Vowel type was alternated across test trials; the pitch value presented first (360 or 450 Hz) was counterbalanced across participants.

Results

As in previous experiments, MLT for each vowel type (360 vs. 450 Hz) was calculated without the first test trial, and these were then submitted to a vowel type (450 Hz vs. 360 Hz) by test trial order (450 Hz first vs. 360 Hz first) ANOVA (see Figure 4). There was no main effect of vowel type, $F(1, 16) = .060$, $p = .810$, no main effect of test order, $F(1, 16) = .521$, $p = .481$, or two-way interaction, $F(1, 16) = .133$, $p = .720$.

Discussion

In Experiment 3, infants showed no listening preference for higher pitch vowels falling within the typical range of an infant talker, indicating that infants do not display a uniform bias favoring higher voice pitch. Rather, based on Experiment 2 it appears that infants are attracted to vowels with infant-like voice pitch values, even when the formant frequencies specify an adult.

Experiment 4

The findings from Experiments 1–3 show that infants are attracted to infant vowels, and that infant-like voice pitch values support this bias. This outcome is consistent with predictions based on infant response to IDS as well as the 'output biases intake' perspective. In Experiment 4, we examined whether infants prefer vowels with infant versus adult formant values when both vowels have voice pitch of an adult female. If infants show a preference here, it would show that the resonance properties (formants) of the infant vowels also contribute to the infant preference for infant vowels. This outcome is not predicted on the basis of infants' responsiveness to IDS, but would be consistent with the 'output biases intake' hypothesis.

Methods

Participants

Data were analyzed from 22 infants, aged 3 to 6 months (10 male, M = 198 days, R = 122–213 days). One additional participant was tested but removed due to MLTs that are outliers with respect to the group mean (1).

Stimuli

The stimuli were four isolated /i/ vowels synthesized using the same parameters as the infant and adult female vowels in Experiment 1, except that now the two infant tokens and the two adult tokens had the same adult female-like pitch values – 210 Hz and 240 Hz. Otherwise, the stimulus files were the same as in Experiment 1.

Procedure and design

The procedure and design were the same as in Experiment 1.

Results

MLT for each vowel type was calculated without the first test trial and then submitted to a vowel type (infant vs. adult) by test trial order (infant first vs. female first) ANOVA (see Figure 4). There was a main effect of vowel type, $F(1, 20) = 4.897$, $p = .039$, $\eta_p^2 = .197$, but no main effect of test order, $F(1, 20) = 0.468$, $p = .502$, or two-way interaction, $F(1, 20) = .945$, $p = .343$.

MLT scores were also compared across Experiments 1 and 4 in which identical adult vowels were presented to infants. MLTs were submitted to a two-way mixed ANOVA with experiment (1 vs. 4) as a between-subjects factor, and vowel type (adult vs. infant) as a within-subjects factor. There were significant main effects for experiment, $F(1, 38) = 8.072$, $p = .007$, $\eta_p^2 = .175$ and vowel type, $F(1, 38) = 19.228$, $p = .000$, $\eta_p^2 = .336$, but no interaction, $F(1, 38) = 2.614$, $p = .114$. Although the interaction failed to reach statistical significance, the effect size associated with vowel type was substantially higher in Experiment 1 ($\eta_p^2 = .506$) compared to Experiment 4 ($\eta_p^2 = .197$).

Discussion

In Experiment 4, infants preferred listening to vowels with infant rather than adult formants when both vowels had voice pitch values appropriate for an adult female. This shows that infants prefer vowels with formant values of a smaller vocal tract. Effect size differences indicate that the listening preference was stronger when the vowels had infant-like voice pitch *and* formant values (Experiment 1) compared to infant-like formants alone (Experiment 4), although cross-experiment comparisons failed to support this pattern as an interaction.

General discussion

In the present study, we examined whether pre-babbling infants preferentially attend to vowels with infant vocal properties over vowels with adult vocal properties. The findings provide some of the first insights into how infants perceive infant speech. Across four experiments, infants listened longer to a vowel with at least one infant vocal property (voice pitch and/or formants) over a vowel with

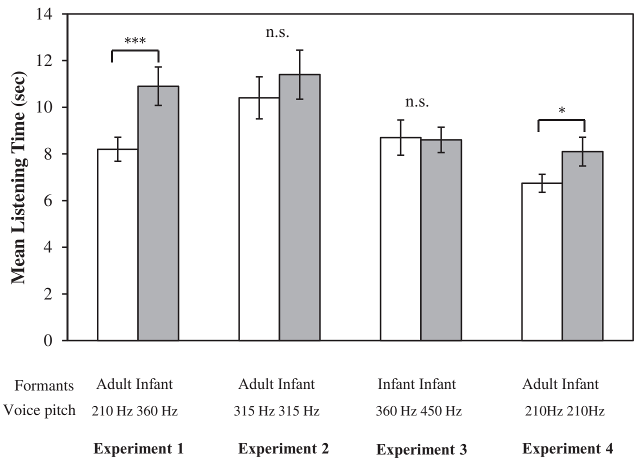

Figure 4 *Mean listening times (sec) to vowel pairs presented in Experiments 1-4. Error bars = std. errors, * p < .05, ** p < .01, *** p < .001.*

no infant-like properties. No preference emerged when both vowels had infant-like voice pitch values. Together, these findings provide the first evidence that young infants display a perceptual bias favoring vowels that have the vocal characteristics of an infant talker.

The preference for infant voice pitch reported here is in line with previous research on infant response to IDS. This finding has theoretical implications for the role of IDS in infant development focusing on the parallels between IDS and infant speech. These parallels support an interpretation of IDS as phonetic convergence, i.e. implicit speech adaptations that mirror features of the interlocutor's speech. Phonetic convergence in caregiver–infant interaction likely conveys positive regard as it does in adult-to-adult interactions (Pardo, 2013). In addition, exposure to infant-like pitch values in IDS may prime the infant's perceptual system for perceiving their *own* speech during vocal learning.

The preference for *both* infant voice pitch and formants observed here is also consistent with the 'output biases intake' view proposed by Vihman and colleagues and provides the first evidence supporting this view in *pre-babbling* infants. To confirm this view, however, future research must determine whether this perceptual bias is present in even younger infants (at 1–3 months) who are not yet able to produce sounds that resemble vowels. It will also be informative to examine whether infants prefer infant speech (e.g. a vowel or syllable) over infant non-speech vocalizations. Such findings will help to determine whether this bias is phonetic in nature or arises from salient acoustic properties common to both infant speech and non-speech vocalizations.

Our ongoing work will build on this new discovery to identify the mechanisms underlying this perceptual bias and to understand how infants may exploit this bias during speech development. A recent study in our lab revealed that infants who were more attentive to infant vowels were also better at recognizing the same vowel produced by different talkers (Polka *et al.*, 2014). This finding suggests that a bias favoring infant speech impacts how infants engage with talker variability, which promotes the formation of speech sound categories (e.g. Rost & McMurray, 2010). Although the present findings do not show how infants perceive their own speech, the perceptual bias we have identified may support them in doing so. For example, an attraction to infant speech could help infants focus attention on their own vocalizations and/or motivate infants to be more vocally active, which in turn may forge perception–action linkages for speech. Future research exploring the underpinnings and impact of this bias will advance our understanding of infant development.

Acknowledgements

This work was supported by Discovery Grants from the Natural Sciences and Engineering Research Council of Canada awarded to L. Polka (DG 105397) and L. Ménard (DG 312395). We thank Leanne Ma and Anush Nersisyan for assistance with infant recruitment and testing, and Robin Panneton, Susan Rvachew, and Catherine Best for feedback on this manuscript.

References

Alhaidary, A, & Rvachew, S. (2011). Babbling and ambient language input: crosslinguistic comparison of corner vowels of 10–18-month-old French and Arabic infants. Poster presented at Society for Research in Child Development Biennial Meeting, 31 March–2 April.

Assman, P.F., & Katz, W.F. (2000). Time-varying spectral change in the vowels of children and adults. *Journal of the Acoustical Society of America*, **108** (4), 1856–1866.

Boersma, P., & Weenink, D. (2013). Praat: doing phonetics by computer (Version 5.3.51) [Computer program]. http://www.praat.org/

Cohen, L.B., Atkinson, D.J., & Chaput, H.H. (2000). *Habit X: a new program for training and organizing data in infant perception and cognition studies (Version 1.0)*. Austin, TX: University of Texas.

Cooper, R.P., & Aslin, R.N. (1994). Developmental differences in infant attention to the spectral properties of infant-directed speech. *Child Development*, **65** (6), 1663–1677.

Danielson, D.K., Seidl, A., Onishi, K.H., Alamian, G., & Cristia, A. (2014). The acoustic properties of bilingual infant-directed speech. *Journal of the Acoustical Society of America Express Letters*, **135** (2), 95–101.

de Boysson-Bardies, B., Hallé, P., Sagart, L., & Durand, C. (1989). A cross-linguistic investigation of vowel formants in babbling. *Journal of Child Language*, **16**, 1–17.

DePaolis, R.A., Vihman, M.M., & Keren-Portnoy, T. (2011). Do production patterns influence the processing of speech in prelinguistic infants? *Infant Behavior and Development*, **34**, 590–601.

DePaolis, R.A., Vihman, M., & Nakai, S. (2013). The influence of babbling patterns on the processing of speech. *Infant Behavior and Development*, **36**, 642–649.

Doupe, A.J., & Kuhl, P.K. (2008). Birdsong and human speech: common themes and mechanisms. In H.P. Zeigler & P. Marler (Eds.), *Neuroscience of birdsong* (pp. 5–31). Cambridge: Cambridge University Press.

Fernald, A., & Kuhl, P.K. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, **10**, 270–293.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B. *et al.* (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, **16**, 477–501.

Gilbert, H.R., Robb, M.P., & Chen, Y. (1997). Formant frequency development — 15 to 36 months. *Journal of Voice*, **11**, 260–266.

Hillenbrand, J., Getty, L.A., Clark, M.J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, **97** (5), 3099–3111.

Kent, R.D. (1992). The biology of phonological development. In C.A. Ferguson, L. Menn & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 65–90). Timonium, MD: York Press.

Kent, R.D., & Murray, A.D. (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *Journal of the Acoustical Society of America*, **72** (2), 353–365.

Kuhl, P.K., Andruski, J.E., Chistovich, I.A., Chistovich, L.A., Kozhevnikova, E.V. *et al.* (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, **277**, 684–686.

Kuhl, P.K., & Meltzoff, A.N. (1996). Infant vocalizations in response to speech: vocal imitation and developmental change. *Journal of the Acoustical Society of America*, **100**, 425–438.

Lam, C., & Kitamura, C. (2012). Mommy, speak clearly: induced hearing loss shapes vowel hyperarticulation. *Developmental Science*, **15**, 212–221.

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America*, **105**, 1455–1468.

Lindblom, B. (1986). Phonetic universals in vowel systems. In J.J. Ohala & J.J. Jaeger (Eds.), *Experimental phonology* (pp. 13–44). New York: Academic Press.

Liu, H.M., Kuhl, P.K., & Tsao, F.M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science*, **6**, F1–F10.

Majorano, M., Vihman, M., & DePaolis, R. (2013). The relationship between infants' production experience and their processing of speech. *Language Learning and Development*, doi:10.1080/15475441.2013.829740

Marie, C., & Trainor, L.J. (2012). Development of simultaneous pitch encoding: infants show a high voice superiority effect. *Cerebral Cortex*, **23**, 660–669. doi:10.1093/cercor/bhs050

Masataka, N. (2003). *The onset of language*. Cambridge: Cambridge University Press.

Ménard, L., Schwartz, J.-L., & Boe, L.-J. (2004). The role of vocal tract morphology in speech development: perceptual targets and sensori-motor maps for French synthesized vowels from birth to adulthood. *Journal of Speech, Language, and Hearing Research*, **47**, 1059–1080.

Pardo, J. (2013). Measuring phonetic convergence in speech production. *Frontiers in Psychology*, **4**, 559. doi:10.3389/fpsyg.2013.00559

Pawlby, S.J. (1977). Imitative interaction. In H.R. Schaffer (Ed.), *Studies in mother–infant interaction* (pp. 203–223). London: Academic Press.

Perry, T.L., Ohde, R.N., & Ashmead, D.H. (2001). The acoustic bases for gender identification from children's voices. *Journal of the Acoustical Society of America*, **109** (6), 2988–2998.

Polka, L., Jusczyk, P.J., & Rvachew, S. (1995). Methods for studying speech perception in infants and children. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language speech research* (pp. 49–89). Timonium, MD: York Press.

Polka, L., Masapollo, M., & Ménard, L. (2014). Who's talking now? Infants' perception of vowels with infant vocal properties. *Psychological Science*, **25**, 1448–1456.

Polka, L., Rvachew, S., & Molnar, M. (2008). Speech perception by 6- to 8-month-olds in the presence of distracting sound. *Infancy*, **13** (5), 421–439.

Rost, G.C., & McMurray, B. (2010). Finding the signal by adding noise: the role of non-contrastive phonetic variability in early word learning. *Infancy*, **15** (6), 608–635.

Rvachew, S., Mattock, K., Polka, L., & Ménard, L. (2006). Developmental and cross-linguistic variation in the infant vowel space: the case of Canadian English and Canadian French. *Journal of the Acoustical Society of America*, **120** (4), 2250–2259.

Rvachew, S., Slawinski, E.B., Williams, M., & Green, C. (1996). Formant frequencies of vowels produced by infants with and without early onset otitis media. *Canadian Acoustics*, **24**, 19–28.

Schroender, M., Atal, B., & Hall, J. (1979). Optimizing digital speech coders by exploiting masking properties of the human ear. *Journal of the Acoustical Society of America*, **66**, 1647–1652.

Shultz, S., & Vouloumanos, A. (2010). Three-month-olds prefer speech to other naturally occurring signals. *Language Learning and Development*, **6**, 241–257.

Strange, W. (1999). Perception of vowels: dynamic constancy. In J.M. Pickett (Ed.), *The acoustics of speech communication* (pp. 153–165). Boston, MA: Allyn & Bacon.

Vorperian, H.K., & Kent, R.D. (2007). Vowel acoustic space development in children: a synthesis of acoustic and

anatomic data. *Journal of Speech, Language, and Hearing Research*, **50**, 1510–1545.
Vouloumanos, A., Hauser, M.D., Werker, J.F., & Martin, A. (2010). The tuning of human neonates' preference for speech. *Child Development*, **81**, 517–527.
Vouloumanos, A., & Werker, J.F. (2004). Tuned to the signal: the privileged status of speech for young infants. *Developmental Science*, **7**, 270–276.
Zhang, Y., Koerner, T., Miller, S., Grice-Patil, Z., Svec, A. *et al.* (2011). Neural encoding of formant-exaggerated speech in the infant brain. *Developmental Science*, **14** (3), 566–581.

Received: 16 February 2014
Accepted: 12 December 2014

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Supplemental Information: When infants talk, infants listen: pre-babbling infants prefer listening to speech with infant vocal properties.