INFANCY
THE OFFICIAL JOURNAL OF THE
INTERNATIONAL CONGRESS
OF INFANT STUDIES

Infancy, 1–12, 2017

ISSN: 1525-0008 print / 1532-7078 online
DOI: 10.1111/infa.12190

WILEY Blackwell

Six-Month-Old Infants Prefer Speech with Raised Formant Frequencies

Titia Benders
Center for Language Studies
Radboud University Nijmegen

Voices can be characterized by their fundamental frequency as well as by speech timbre characteristics, such as formant frequencies. While infants' responses to fundamental frequency characteristics are well established, very little is known about how infants respond to changes in formant frequencies. This study tested whether 6-month-old infants prefer listening to speech with raised formant frequencies over speech with lowered formant frequencies. Naturally spoken utterances were acoustically manipulated to render raised-formant and lowered-formant stimuli that only differed in the formant frequencies, while keeping fundamental frequency and other acoustic characteristics constant. Infants in an infant-controlled listening procedure listened longer to the raised-formant than to the lowered-formant stimuli. These results provide the first evidence that infants distinguish and show preferences for variations in formant frequencies in adult speech.

Infants employ all their senses to navigate the multimodal emotional landscape they are born into (for a review, Grossmann, 2010). Infants preferably attend to stimuli that convey positive affect (Kim & Johnson, 2013) and may perceive affective information more strongly from voices than from faces (Grossmann, 2010; Vaish, Grossmann, & Woodward, 2008). It thus appears inevitable that infants prefer happy over sad speech (Singh, Morgan, & Best, 2002). Caregivers provide their infants with such positive vocal stimuli in the form of infant-directed speech (IDS; see Soderstrom, 2007). To further understand infants' speech preferences, it is critical to identify the speech characteristics that guide them.

Research into IDS has long recognized that caregivers convey affect through the use of fundamental frequency (F0) contours; for example, approval is cued by a rising F0, while prohibition is cued by a rapidly falling F0 (Fernald, 1992; Papoušek, Papoušek, & Symmes, 1991). Infants of 4 and 5 months old already prefer listening to F0 contours that are associated with positive affect over contours that express negative affect (Fernald, 1993; Papoušek, Bornstein, Nuzzo, Papoušek, & Symmes, 1990).

Correspondence should be sent to Titia Benders, Department of Linguistics, Macquarie University, Australian Hearing Hub, Level 3 North, Room 435, Macquarie University, NSW 2109, Australia. E-mail: titia.benders@mq.edu.au

Research into the vocal expression of emotion in non-infant-directed speech has identified another vocal cue to affect, namely speech timbre, or the composition of the speech spectrum (Scherer, 1986). A natural speech signal has harmonics at multiples of the F0. The shape of the vocal tract reinforces sound energy in particular regions of this frequency spectrum. The resulting spectral peaks are called formants (Fant, 1960). The formants are one aspect of speech timbre.

As different emotions—for example, joy and sadness—are expressed with specific shapes of the vocal tract—for example, smiles versus frowns—the facial expression of emotion also affects formant frequencies. Specifically, smiled speech has higher formant frequencies than frowned speech (Tartter & Braun, 1994). It is perhaps unsurprising that IDS, which is accompanied by many smiles (Chong, Werker, Russell, & Caroll, 2003), also has overall higher formant frequencies than adult-directed speech (ADS; Englund & Behne, 2005; Green, Nip, Wilson, Mefferd, & Yunusova, 2010; Benders, 2013).

Formant frequencies are not just an acoustic correlate of smiles and frowns, they also enable adult listeners to recognize and interpret a speaker's affective state (Scherer, 2003; Tartter & Braun, 1994). It is so far unknown whether formant frequencies have any effect on the recognition of vocal emotion by infants. This study is a first step toward addressing this issue, by asking whether infants prefer listening to speech with raised formant frequencies over speech with lowered formant frequencies.

Infants from a very young age are clearly sensitive to formant frequencies, as even newborns are able to discriminate between vowels (Aldridge, Stillman, & Bower, 2001). Vowels such as /i/ and /u/ sound different from each other because they have distinct formant patterns. There is evidence that the spectral properties of speech also contribute to infants' speech preferences, which comes from studies showing that the F0 presented without the harmonics may not elicit a preference for IDS (Colombo & Horowitz, 1986; Kaplan, Goldstein, Huckeby, Owren, & Cooper, 1995; Panneton Cooper & Aslin, 1994; but see Fernald & Kuhl, 1987). However, harmonics are not sufficient to elicit the preference, as infants do not prefer IDS over ADS when the harmonics are presented without the F0 (Kaplan et al., 1995). Building on the knowledge that infants' speech preferences clearly profit from the presence of spectral information, this study asks whether infants' speech preferences are also guided by the exact spectral composition of the speech, specifically by the raising or lowering of the formant frequencies.

Direct evidence that infants' attention is attracted to higher formant frequencies is provided by studies showing that 4- to 5-month-olds prefer infant-produced vowels over adult-produced vowels (Masapollo, Polka, & Ménard, 2016). Infants have smaller vocal folds and vocal tracts than adults, and infant-produced vowels consequently are higher pitched and have higher formant frequencies than adult-produced vowels (for a review, see Vorperian & Kent, 2007). The higher formant frequencies independently contribute to infants' preference for infant-produced vowels (Masapollo et al., 2016). However, this preference has been established with isolated vowel stimuli and with formant frequencies that extend beyond the range of what an adult vocal tract typically produces. It is an open question whether infants' preference for speech with high formant frequencies extends to connected speech with formant frequencies that could have been produced by an adult voice.

This study is the first to examine whether infants prefer adult speech with raised formant frequencies over adult speech with lowered formant frequencies. This preference is assessed in infants at 6 months of age, because 6-month-old infants prefer happy over neutral speech in both IDS and ADS (Singh et al., 2002) and are more interested in approving speech than younger or older infants (Kitamura & Lam, 2009).

METHODS

Tightly controlled auditory stimuli, which only differed in their formant structure, were presented to 6-month-old infants in an infant-controlled head-turn preference procedure. Longer listening times to the stimuli in the raised-formant condition than to the stimuli in the lowered-formant condition would provide evidence that infants' preferences are driven by formant frequency shifts.

Participants

A total of 30 6-month-old infants participated in this experiment (15 male, 15 female; mean age: 199 days; age range: 180–216 days). The data from 23 participants were included in the data analysis (13 male, 10 female; mean age: 199 days; age range: 180–216 days). Seven participants were excluded from the analyses because they were ill at the time of testing ($n = 1$); cried during the experiment ($n = 3$); were too restless to complete the experiment ($n = 2$); or their parent interfered with the experiment ($n = 1$). None of the participants had an acute or chronic ear infection.

The number of participants to be tested (30) was determined a priori based on experience with the dropout rate in the head-turn preference procedure and a power analysis suggesting that a sample of 21 was required for a power of at least .8. A full description of the procedure that was followed to determine the sample size can be found in the Appendix S1.

Participants were recruited from a subject database that is maintained at the Baby Research Center at the Radboud University Nijmegen. Infants were invited to participate if they were born full term, had no familial risk of language or reading problems, and were from families in which Dutch was (one of) the primary language(s). Parents of participants received an age-appropriate book for their child or a small monetary compensation (€10.00) for their participation.

Stimuli—individual sentences

The sentences for the stimulus materials were taken from an audiology test developed by Versfeld, Daalder, Festen, and Houtgast (1999). That test contains several sets of 13 sentences ("Sentence Sets"). Sentences are eight or nine syllables long, with words of up to three syllables long. Each Sentence Set has a frequency distribution of phonemes that closely matches the distribution in Dutch. Each trial in the infant experiment corresponded to one Sentence Set.

A female speaker from the Nijmegen area with a clear voice quality recorded Sentence Sets 1 through 5, by repeating these sentences from the original (Versfeld et al., 1999) recordings.[1] This recording procedure ensured that the materials were spoken with a neutral intonation and at a constant pace.

---

[1] Specifically, our speaker repeated the sentences as spoken by female speaker HB.

The raised-formant and lowered-formant versions of the recorded sentences were created via a manipulation of the sampling frequency in the computer program Praat (Boersma & Weenink, 2010). The algorithm first overrides the sampling frequency of the original sound with a higher sampling frequency to effectively raise the spectrum or with a lower sampling frequency to effectively lower the spectrum and then adds the original pitch and duration back to this manipulated signal using the overlap-add method (Moulines & Charpentier, 1990). This procedure resulted in sentences with the same duration and pitch characteristics as the original recordings, but a shift to all formant frequencies.

Six adult native speakers of Dutch in a first control experiment adapted the sentences from Sentence Set 1 in an interactive procedure to create “happy-sounding” or “sad-sounding” versions of the sentences, all the while ensuring that the resulting sentences still sounded natural. Listeners created the “happy-sounding” versions with a formant shift ratio of, on average, 1.13 and created the “sad-sounding” versions with a formant shift ratio of, on average, 0.87. These average shift factors were employed to create, respectively, the raised-formant stimuli and the lowered-formant stimuli for the experiment.

Each recorded sentence was manipulated twice, to render a raised-formant and a lowered-formant version, and then scaled to equate average intensity across all sentences. See Figure 1 for visual representations of example stimuli. Note that the conditions only differed in their formant frequencies (subfigures c), as the amplitude and fundamental frequency of the two versions of each sentence were identical (subfigures a and b).

Eight native speakers of Dutch in a second control experiment rated all stimuli on three continuous scales: facial expression (“frown” to “smile”); energy level (“tired” to “energetic”); and naturalness of the voice (“very unnatural” to “fully natural”). The raised-formant versions of the sentences were judged to sound more smiley and more energetic than the lowered-formant versions of the sentences. The naturalness of the sentences was rated to be “somewhat natural” and, most importantly, highly similar between the conditions. A full description of this second control experiment can be found in the Appendix S1.

Stimuli—sentence sequences

A stimulus in the infant experiment was a 30-sec sequence of either the raised-formant or the lowered-formant sentences from one Sentence Set.

The familiarization stimuli were created from Sentence Set 1. The raised-formant familiarization sequence contained the sentences in the opposite order from the lowered-formant familiarization sequence.

The test stimuli were created from Sentence Sets 2 through 5. The sentences from each Sentence Set were combined into two sequences per condition, with the second sequence presenting the sentences in the reversed order from the first. This procedure resulted in eight raised-formant sequences and eight lowered-formant sequences. Each participant would hear all 16 sequences in the experiment.

The sentences were concatenated into sequences with the first sentence starting at 250 msec and an offset-to-onset silent interval of 650 msec. The sequences were cut at exactly 30 sec, with a 1-sec fade-out for a less abrupt ending. The resulting 30-sec

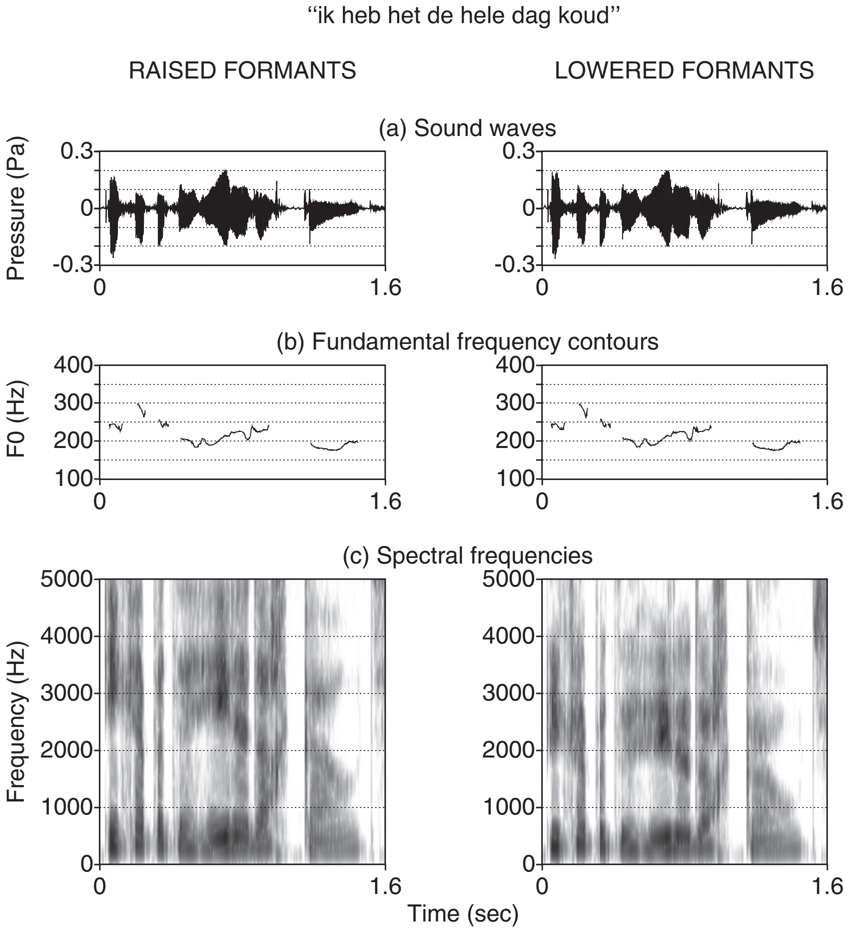

Figure 1 Example stimulus "ik heb het de hele dag koud" as used in the experiment. Panels display the wave form (panel a); the fundamental frequency contour (panel b); and the spectrogram (panel c) in the raised-formant (left) and lowered-formant (right) version.

stimuli consisted of approximately 11 full sentences each. The stimuli were saved in WAVE format for presentation in the experiment.

EQUIPMENT AND PROCEDURE

Infants were tested in a head-turn preference procedure. This has been the most frequently used procedure to test infants' preferences for IDS over ADS (Dunst, Gorman, & Hamby, 2012).

Infants sat on their parent's lap, in the center of a three-sided testing booth. Caregivers wore headphones with salsa music and a superimposed speech stream with various alternating speakers. Caregivers were instructed to limit interaction with their infant to a minimum. The infant was positioned opposite a blue center light in the middle panel of the testing booth, and exactly in between two red side lights in the left and right walls of the booth. All lights were placed at infant eye level. Loudspeakers were located underneath the red side lights, hidden from view by the material of the booth. Stimuli played from the loudspeakers and were calibrated to have a volume of 65 dB(A) at the approximate location of the infant's ears. A camera was hidden beneath the center light and recorded the infant. A curtain behind the parent and the infant closed the testing booth.

The experimenter remained behind that curtain outside of the testing booth, observed the infant via the camera, and used a computer keyboard to code the infant's behavior as looking to the center, left, or right light. The experimenter listened to the same masking music as the caregiver for the duration of the experiment and was blind to the experiment phase (familiarization or test) and the trial condition (raised-formant or lowered-formant).

The experiment was presented and controlled using the Lincoln Infant Lab Package (Meints & Woodford, 2008). This program also recorded the online coding by the experimenter.

Each trial started with the flashing of the blue center light. Once the experimenter indicated that the infant fixated the center light, a trial would start. At this point, the center light stopped flashing and one of the red side lights began to flash. When the experimenter indicated that the infant made a head turn of at least 30 degrees toward the flashing side light, the sound stimulus began playing from the loudspeaker on that side. The trial would end when the experimenter indicated that the infant looked away from the flashing light for two consecutive seconds. If the infant looked away and returned to the light within 2 sec, the trial would continue. The maximum trial duration was 30 sec, after which the trial would end automatically.

The procedure started with a familiarization phase during which the infant needed to accumulate a minimum of 15 sec of listening time to the raised-formant as well as the lowered-formant familiarization trial, to ensure that looks to early trials would not be due to general novelty effects (see Singh et al., 2002, for the same familiarization phase). The familiarization trials alternated between the conditions; the presentation side was pseudo-randomized, with the restriction that the same side did not appear more than twice in a row; and the association between the conditions and presentation sides was counterbalanced within infants. Once the infant had accumulated a minimum of 15 sec of familiarization to a stimulus type, the current familiarization trial would be finished as usual, so as not to affect the infant's understanding of the procedure. When 15 sec of listening time to both stimulus types had been recorded, the experiment would proceed to the test phase.

The test phase consisted of 16 trials, each presenting one of the 16 test stimuli. The 16 trials were divided over two blocks of four raised-formant and four lowered-formant trials. Recall that each Sentence Set rendered two raised-formant and two lowered-formant sequences. The two sequences of a condition that were created from the same Sentence Set only differed in the order of the sentences. The raised-formant stimuli with the first sentence order and the lowered-formant stimuli with the second sentence order were presented in the first block. The remaining stimuli were presented in the second block. The association between the conditions and sides was counterbalanced within infants, and the order of the trials was pseudo-randomized, to ensure that the conditions (raised-formant and lowered-formant) and the presentation sides (left and right) did not appear more than twice in a row.

Reliability coding

Offline coding of the test trials of 10 randomly selected successfully tested participants was performed on a 40-msec frame-by-frame basis by a trained coder using SuperCoder (Hollich, 2005). The coder, who had not been the experimenter, was blind to the trial condition (raised-formant or lowered-formant). The correlation between offline- and online-coded total looking durations on the 16 test trials ranged from .991 to .999 across the 10 participants. The mean difference between offline-coded and online-coded total looking durations ranged from −379 to 337 msec across participants. Online-coded looking times were longer than offline-coded looking times for five of the recoded participants, and shorter for the other five. Compared to the offline coding, the online coding underestimated the total looking durations in the raised-formant condition by 9 msec and overestimated the total looking durations in the lowered-formant condition by 35 msec. These small differences suggest that, if anything, the online coding underestimated the effect of interest. Overall, these results showed a high reliability of the online coding and the results were based on the online-coded looking times only.

Data cleaning and preparation, analysis plan

All infants looked on average longer than 4 sec per trial in the test phase. One trial with a looking time of 97 msec was excluded as displaying a floor effect. All other trials had looking times longer than 1 sec.[2] All trials had looking times below the maximum trial duration of 30 sec, eliminating the need to exclude trials as displaying ceiling effects. One participant contributed 15 trials to the analysis; the remaining 22 participants contributed all 16 trials.

Average looking times in the raised-formant and lowered-formant conditions were compared by means of a paired-samples *t*-test. The test was conducted under the assumption of equal variances ($F(22,22) = 0.816, p = .637$), and results were interpreted at an α level of .05. All data processing and all analyses were conducted in the open-source statistical software R (R Development Core Team, 2004).

RESULTS AND CONCLUSION

The mean looking times per condition as well as the difference between conditions (raised-formant minus lowered-formant) are displayed in Figure 2. Infants looked significantly longer when hearing stimuli with raised formants than stimuli with lowered formants ($t[22] = 2.263, p = .034$, Cohen's $d = 0.479$[3] ; raised-formant condition: $M = 8.209$ sec, $SD = 3.288$; lowered-formant condition: $M = 7.407$ sec, $SD = 2.969$;

---

[2] Previous studies that investigated infants' preference for IDS in the head-turn preference procedure typically excluded looking times below 2 sec as floor effects. However, formant frequencies are present in every vowel as well as some consonants. As the longest first syllable across all trials was 188 msec and trials started with 250 ms silence, at least one syllable had thus been presented 438 msec into each trial. Looks away after 1 sec into the trial could thus have been responses to the formant frequencies of the stimulus and were retained in the analysis.

[3] Corrected for the correlation between the two conditions using equation 8 in Morris and DeShon (2002). Correlation between conditions: $r(21) = .857$

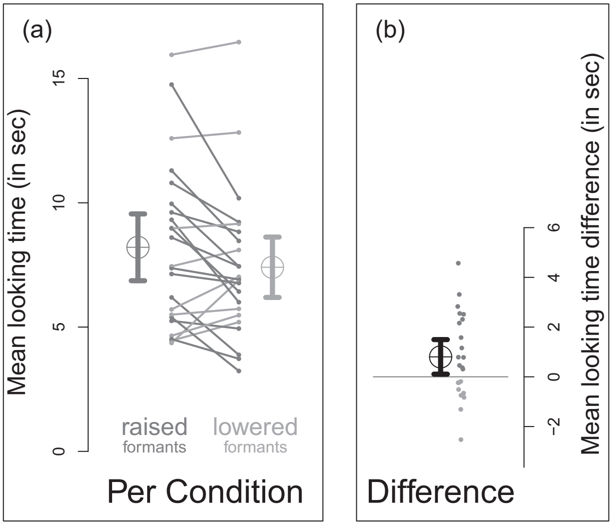

Figure 2 (Panel a) Looking times in the raised-formant condition (dark gray) and the lowered-formant condition (light gray). Open circles with error bars indicate the mean and its 95% (between-subjects) confidence interval for each condition. Closed circles give individual infants' mean looking times, which are connected between the conditions by lines that indicate whether the individual infant looked longer in the raised-formant conditions (dark gray) or in the lowered-formant condition (light gray). (Panel b) The mean looking time difference between conditions (raised-formant minus lowered-formant condition). The open circle with error bars indicates the mean and its 95% (within-subjects) confidence interval. The closed circles give individual infants' mean difference, with differences above 0 indicating longer looking times in the raised-formant condition (dark gray) and differences below 0 indicating longer looking times in the lowered-formant condition (light gray).

difference M = 0.802 sec, difference SD = 1.700). This result shows that infants prefer speech with raised formant frequencies over speech with lowered formant frequencies.

DISCUSSION

The present study examined whether 6-month-old infants preferably attend to adult speech with raised formant frequencies over adult speech with lowered formant frequencies. The results show that they do.

The procedure employed here establishes a preference, but does not provide direct insight into what guides the preference. The observed preference could be a familiarity preference, as speech with higher formant frequencies may be prevalent in infants' input through infant-directed speech (IDS; Englund & Behne, 2005; Green et al., 2010; Benders, 2013), more female than male speech (Peterson & Barney, 1952), or infants' own emerging productions that they may be listening to (Peterson & Barney, 1952; DePaolis, Vihman, & Keren-Portnoy, 2011; DePaolis, Vihman, & Nakai, 2013; Majorano, Vihman, & DePaolis, 2014).

The literature also suggests two richer accounts of why infants might prefer speech with raised formant frequencies, which go beyond a familiarity preference. Firstly, the preference for raised formant frequencies may be an instantiation of infants' general preference for happy vocal output (Corbeil, Trehub, & Peretz, 2013; Singh et al., 2002). Infants may have learned that high formant frequencies often coincide with happy speech by listening to adults who smile or use IDS (Benders, 2013; Englund & Behne, 2005; Green et al., 2010; Tartter & Braun, 1994). Secondly, infants may prefer speech from physically smaller speakers over speech from larger speakers. In that sense, the present findings could be a variant on the observation that infants prefer listening to the high formant frequencies from infant-produced vowels (Masapollo et al., 2016).

The present study cannot disentangle these two accounts of infants' preference for raised formant frequencies by virtue of employing a preference procedure. Moreover, formant frequencies are inherently associated with a multitude of factors, and the affect-based and size-based accounts of infants' preference for raised formant frequencies may be intertwined: It has been hypothesized that the smile is a signal of positive affect exactly because it makes the speaker sound smaller (Ohala, 1984).

The affect-based interpretation of the present findings is specifically limited by the lack of a unique association between raised formant frequencies and happy speech. Raised formant frequencies also occur in angry speech (Scherer, 2003; Waaramaa, Laukkanen, & Väyrynen, 2008), possibly because lip retraction also conveys dominance (Niedenthal, Mermillod, Maringer, & Hess, 2010). Furthermore, raised formant frequencies are one of several timbre cues associated with affect that may guide infants' speech preferences. Neither of these limitations to the affect-based interpretation are unique to the present study or to formant frequencies. For example, the high fundamental frequency and large fundamental frequency range may contribute to infants' preference for infant-directed speech (Fernald & Kuhl, 1987), but occur in angry speech as well (Scherer, 2003). Moreover, most previous work on infants' speech preferences has focused on fundamental frequency cues, without attention to any aspect of timbre, including the presently addressed formant frequencies. The present study thus constitutes the next step to identifying the acoustic cues that guide infants' speech preferences.

A final limitation to the study is that only the effect of one relatively large degree of difference between raised and lowered formant frequencies was tested. The ratio between formant frequencies in IDS and adult-directed speech (ADS) appears to be around 1.15 (based on figures in Englund & Behne, 2005; Green et al., 2010; Benders, 2013), which compares well to the ratio of 1.13 that was employed to create the raised-formant stimuli from the neutrally sounding recordings, but is larger than the overall ratio of 1.30 between the raised-formant and lowered-formant stimuli. The 1.30 ratio between the stimuli is also considerably larger than the ratio of around 1.07 between smiled and frowned adult-directed speech (based on tables in Tartter & Braun, 1994). Nevertheless, the effect size observed in the present study, Cohen's $d = 0.479$, falls below the Cohen's $d = 0.57$ lower bound of the 95% confidence interval of infants' preference for IDS over ADS Dunst et al., 2012). This suggests that infants' preference for raised over lowered formant frequencies may contribute relatively little to their preference for, for example, IDS.

Despite the multiple explanations and limitations, the present results advance our understanding of the potential positive outcomes of raised formant frequencies in IDS (Benders, 2013; Englund & Behne, 2005; Green et al., 2010). IDS has been attributed three functions: directing attention, expressing affect, and teaching the infant language (Fernald, 1992). The observed preference for raised formant frequencies suggests that they contribute to the attention-grabbing qualities of IDS. Moreover, any characteristic that draws infants' attention to the speech signal may indirectly contribute to infants' language learning (Graf Estes & Hurley, 2013; cf. Song, Demuth, & Morgan, 2010). Note that the latter proposal is crucially different from the hypothesis that enhanced formant-frequency contrasts between the corner vowels in IDS specifically facilitate vowel learning (Burnham, Kitamura, & Vollmer-Conna, 2002; De Boer & Kuhl, 2003; Escudero, Benders, & Wanrooij, 2011; Kuhl et al., 1997; Liu, Kuhl, & Tsao, 2003; Uther, Knoll, & Burnham, 2007; cf. Benders, 2013; Englund & Behne, 2005; Martin et al., 2015). Rather, raised formant frequencies may contribute to language learning more generally.

The present results also provide new avenues for understanding how infants learn to match common affect across voices and faces (Grossmann, Striano, & Friederici, 2006; Kahana-Kalman & Walker-Andrews, 2001; Soken & Pick, 1999), which may initially "be based on learning to associate a certain facial expression with the vocal expression that consistently accompanies it" (Grossmann, 2010, p. 231). Infants under 7 months of age crucially rely on information from the mouth when matching faces to emotional voices (Walker-Andrews, 1986). As mouth shapes consistently affect formant frequencies, but for example not the fundamental frequency, it appears worth testing whether the consistent co-occurrence of mouth shapes and formant frequencies scaffolds infants into matching affect across faces and voices more generally. The present findings lay the foundations for such work, by showing that infants are sensitive to variations in formant frequencies.

ACKNOWLEDGMENTS

The author would like to acknowledge Irene Strikkers, Leanne Nagels, and Imme Lammertink for their help in stimulus preparation. Thanks go to the public library in Nijmegen, for making available a testing room for the pretests, and to all the volunteer adult listeners that lent their time and ears to this research. Leanne Nagels has been invaluable in testing the infant participants. This research would not have been possible without the contribution of the infant participants and their parents. Earlier drafts of this manuscript were improved with the help of comments from the members of the First-Language Acquisition group at the Radboud University Nijmegen.

REFERENCES

Aldridge, M., Stillman, R., & Bower, T. (2001). Newborn categorization of vowel-like sounds. *Developmental Science*, *4*(2), 220–232.

Benders, T. (2013). Mommy is only happy! Dutch mothers realisation of speech sounds in infant-directed speech expresses emotion, not didactic intent. *Infant Behavior and Development*, *36*, 847–862.

Boersma, P., & Weenink, D. (2010).*Praat, doing phonetics by computer*. (Version 5.3.46). Retrieved from http://www.praat.org on 22 April 2013

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science*, *296*, 1435.

Chong, S., Werker, J., Russell, J., & Caroll, J. (2003). Three facial expressions mothers direct to their infants. *Infant and Child Development*, *12*, 211–232.

Colombo, J., & Horowitz, F. D. (1986). Infants' attentional responses to frequency modulated sweeps. *Child Development*, *57*(2), 287–291.

Corbeil, M., Trehub, S., & Peretz, I. (2013). Speech vs. singing: Infants choose happier sounds. *Frontiers in Psychology*, *4*, 372.

De Boer, B., & Kuhl, P. (2003). Investigating the role of infant-directed speech with a computer model. *Acoustics Research Letters Online*, *4*(4), 129–134.

DePaolis, R., Vihman, M., & Keren-Portnoy, T. (2011). Do production patterns influence the processing of speech in prelinguistic infants? *Infant Behavior and Development*, *34*(4), 590–601.

DePaolis, R., Vihman, M., & Nakai, S. (2013). The influence of babbling patterns on the processing of speech. *Infant Behavior and Development*, *36*(4), 642–649.

Dunst, C. J., Gorman, E., & Hamby, D. W. (2012). Preference for infant-directed speech in preverbal young children. *Center for Early Literacy Learning*, *5*(1), 1–13.

Englund, K., & Behne, D. (2005). Infant directed speech in natural interaction – Norwegian vowel quantity and quality. *Journal of Psycholinguistic Research*, *34*(3), 259–280.

Escudero, P., Benders, T., & Wanrooij, K. (2011). Enhanced bimodal distributions facilitate the learning of second language vowels. *Journal of the Acoustical Society of America*, *130*(4), EL206–EL212.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Fernald, A. (1992). Meaningful melodies in mothers' speech to infants. In H. Papousek, U. Jurgens, & M. Papousek (Eds.), *Origins and development of nonverbal vocal communciation: Evolutionary, comparative, and methodological aspects* (p. 262). Cambridge: Cambridge University Press.

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.

Graf Estes, K., & Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meanings. *Infancy*, *18*(5), 797–824.

Green, J., Nip, I., Wilson, E., Mefferd, A., & Yunusova, Y. (2010). Lip movement exaggerations during infant-directed speech. *Journal of Speech, Language, and Hearing Research*, *53*, 1529–1542.

Grossmann, T. (2010). The development of emotion perception in face and voice during infancy. *Restorative Neurology and Neuroscience*, *28*, 219–236.

Hollich, G. (2005). Supercoder: A program for coding preferential looking. [Computer Software]. West Lafayette: Purdue University.

Grossmann, T., Striano, T., & Friederici, A. D. (2006). Crossmodal integration of emotional information from face and voice in the infant brain. *Developmental Science*, *9*(3), 309–315.

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

Kaplan, P. S., Goldstein, M. H., Huckeby, E. R., Owren, M. J., & Cooper, R. P. (1995). Dishabituation of visual attention by infant-versus adult-directed speech: Effects of frequency modulation and spectral composition. *Infant Behavior and Development*, *18*(2), 209–223.

Kim, H. I., & Johnson, S. P. (2013). Do young infants prefer an infant-directed face or a happy face? *International Journal of Behavioral Development*, *37*(2), 125–130.

Kitamura, C., & Lam, C. (2009). Age-specific preferences for infant-directed affective intent. *Infancy*, *14*(1), 77–100.

Kuhl, P., Andruski, J., Chistovich, I., Chistovich, L., Kozhevnikova, E., Ruskina, V., … Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, *277*(5044), 684–686.

Liu, H.-M., Kuhl, P., & Tsao, F.-M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science*, *6*(3), F1–F10.

Majorano, M., Vihman, M., & DePaolis, R. (2014). The relationship between infants production experience and their processing of speech. *Language Learning and Development*, *10*(2), 179–204.

Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., & Cristia, A. (2015). Mothers speak less clearly to infants than to adults a comprehensive test of the hyperarticulation hypothesis. *Psychological Science*, *26*(3), 341–347.

Masapollo, M., Polka, L., & Ménard, L. (2016). When infants talk, infants listen: Pre-babbling infants prefer listening to speech with infant vocal properties. *Developmental Science*, *19*, 318–328.

Meints, K., & Woodford, A. (2008). *Lincoln infant lab package 1.0: A new programme package for ipl, preferential listening, habituation and eyetracking [Computer software manual]*. Retrieved from http://www.lincoln.ac.uk/psychology/babylab.htm

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*(1), 105.

Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, *9*, 453–467.

Niedenthal, P. M., Mermillod, M., Maringer, M., & Hess, U. (2010). The simulation of smiles (sims) model: Embodied simulation and the meaning of facial expression. *Behavioral and Brain Sciences*, *33*(06), 417–433.

Ohala, J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica*, *41*, 1–16.

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

Papoušek, M., Bornstein, M., Nuzzo, C., Papoušek, H., & Symmes, D. (1990). Infant responses to prototypical melodic contours in parental speech. *Infant Behavior and Development*, *13*, 539–545.

Papoušek, M., Papoušek, H., & Symmes, D. (1991). The meanings of melodies in motherese in tone and stress languages. *Infant Behavior and Development*, *14*, 415–440.

Peterson, G., & Barney, H. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, *24*(2), 175–184.

R Development Core Team. (2004). *R: A language and environment for statistical computing [Computer software manual]*. Vienna, Austria. Retrieved from http://www.R-project.org

Scherer, K. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, *99*, 143–165.

Scherer, K. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, *40*, 227–256.

Singh, L., Morgan, J., & Best, C. (2002). Infants' listening preferences: Baby talk or happy talk? *Infancy*, *3* (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.

Soken, N., & Pick, A. (1999). Infants' perception of dynamic affective expressions: Do infants distinguish specific expressions? *Child Development*, *70*(6), 1275–1282.

Song, J. Y., Demuth, K., & Morgan, J. (2010). Effects of the acoustic properties of infant-directed speech on infant word recognitiona. *The Journal of the Acoustical Society of America*, *128*(1), 389–400.

Tartter, V., & Braun, D. (1994). Hearing smiles and frowns in normal and whisper registers. *Journal of the Acoustical Society of America*, *96*(4), 2101–2107.

Uther, M., Knoll, M., & Burnham, D. (2007). Do you speak E-NG-L-I-SH? A comparison of foreigner-and infant-directed speech. *Speech Communication*, *49*, 2–7.

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin*, *134*(3), 393–403.

Versfeld, N., Daalder, L., Festen, J., & Houtgast, T. (1999). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *Journal of the Acoustical Society of America*, *107*(3), 1671–1684.

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*(6), 1510–1545.

Waaramaa, T., Laukkanen, A.-M., & Väyrynen, E. (2008). Monopitched expression of emotions in different vowels. *Folia Phoniatrica et Logopaedica*, *60*, 249–255.

Walker-Andrews, A. S. (1986). Intermodal perception of expressive behaviors: Relation of eye and voice? *Developmental Psychology*, *22*(3), 373–377.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article:

Appendix S1. Determination of sample size and adult ratings of the stimuli.