Published in final edited form as:

Interact Stud. 2014 ; 15(1): 38–54. doi:10.1075/is.15.1.02smi.

Infant-Directed Visual Prosody: Mothers' Head Movements and Speech Acoustics

Nicholas A. Smith and Heather L. Strader

Boys Town National Research Hospital

Abstract

Acoustical changes in the prosody of mothers' speech to infants are distinct and near universal. However, less is known about the visible properties mothers' infant-directed (ID) speech, and their relation to speech acoustics. Mothers' head movements were tracked as they interacted with their infants using ID speech, and compared to movements accompanying their adult-directed (AD) speech. Movement measures along three dimensions of head translation, and three axes of head rotation were calculated. Overall, more head movement was found for ID than AD speech, suggesting that mothers exaggerate their visual prosody in a manner analogous to the acoustical exaggerations in their speech. Regression analyses examined the relation between changing head position and changing acoustical pitch (F_0) over time. Head movements and voice pitch were more strongly related in ID speech than in AD speech. When these relations were examined across time windows of different durations, stronger relations were observed for shorter time windows (< 5 sec). However, the particular form of these more local relations did not extend or generalize to longer time windows. This suggests that the multimodal correspondences in speech prosody are variable in form, and occur within limited time spans.

One of the most distinctive aspects of mother-infant interaction is the prosody of infant-directed (ID) speech. Acoustical analyses of ID speech have shown that mothers' ID speech has a higher mean pitch, distinctive and expanded pitch contours, longer pauses and increased repetition relative to adult-directed (AD) speech (Fernald & Simon, 1984; Fernald et al., 1989; Stern, Spieker, Barnett, & MacKain, 1983; Stern, Spieker, & MacKain, 1982). Infants are sensitive to this prosody, and they prefer to listen to ID speech over AD speech (Cooper & Aslin, 1990; Fernald, 1985; Pegg, Werker, & McLeod, 1992; Werker, Pegg, & McLeod, 1994), primarily due to its exaggerated pitch properties (Fernald & Kuhl, 1987). Mothers' prosody is, in turn, modulated by feedback from their infants (Smith & Trainor, 2008).

Speech is multisensory because the acoustical speech signal is produced through the movement of articulators that shape the vocal tract, and many of these movements are visible on the lips, face and head (Graf, Cosatto, Strom, & Huang, 2002; Hadar, Steiner, Grant, & Rose, 1983; Yehia, Rubin, & Vatikiotis-Bateson, 1998). Examinations of speech between adults have found that talker head movements are related to suprasegmental

properties of speech, such as rhythm and stress, and help to regulate turn taking (Hadar et al., 1983; Hadar, Steiner, Grant, & Rose, 1984; Hadar, Steiner, & Rose, 1984). Significant correlations have been observed between time varying head movement measures and acoustical measures of F_0 (Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004; Yehia, Kuratate, & Vatikiotis-Bateson, 2002). Even beyond conventional speech, listeners are able to judge the size of musical pitch intervals produced by singers on the basis of visual information alone (Thompson & Russo, 2007). Subjects are able to recognize individual talkers on the basis of correspondence between facial dynamics and speech acoustics in a delayed matching task with videos of unfamiliar faces and the sounds of unfamiliar voices (Kamachi, Hill, Lander, & Vatikiotis-Bateson, 2003). In tests of audiovisual speech perception the intelligibility of speech in noise is greater when the talker's natural head movements are present (Davis & Kim, 2006; Munhall et al., 2004). Altogether these studies examining the visual prosody of adult speech suggest that visual and acoustical prosody of ID speech might also be related.

Although ID speech has been examined in many studies both in terms of maternal production and infant perception, these studies have primarily focused on the acoustical aspects of this multisensory phenomenon. The goal of the present study is to provide an objective description of the accompanying visual prosody of ID speech. The ID quality of communication is not restricted to speech acoustics, but has been shown for ID action, or "motionese" (Brand, Baldwin, & Ashburn, 2002), ID gesture or "gesturese" (O'Neill, Bard, Linnell, & Fluck, 2005) and ID sign language (Masataka, 1998). Recent work has found positive correlations between lip movements and the hyperarticulation of vowels that is characteristic of ID speech (Green, Nip, Wilson, Mefferd, & Yunusova, 2010). The present study extends this work by examining (1) how visual prosody – defined in this study as the expressive head movements that accompany speech – may differ in ID and AD speech, and (2) the relation between visual prosody and acoustical measures of prosody in ID and AD speech.

Infants are sensitive to the audiovisual properties of speech. They can match visual presentations of talking faces with the appropriate auditory presentation of the speech sound (Hollich, Newman, & Jusczyk, 2005; Hollich & Prince, 2009; Kuhl & Meltzoff, 1984; Kuhl & Meltzoff, 1982; Patterson & Werker, 2002), and can discriminate their native from non-native language using amodal properties such as synchrony (Bahrick & Pickens, 1988) or visual-only information (Weikum et al., 2007). Visual information can influence infants' phoneme perception (Rosenblum, Schmuckler, & Johnson, 1997), and enhance phoneme discrimination (Teinonen, Aslin, Alku, & Csibra, 2008).

Beyond the multisensory aspects of speech, intersensory redundancy is an important aspect of mother-infant interaction. In particular, time-locked multimodal experiences provide a powerful mechanism for early word learning (Gogate & Walker-Andrews, 2001; Gogate, Walker-Andrews, & Bahrick, 2001; Smith & Gasser, 2005), and mothers' use of temporal synchrony in their verbal labeling and gestures suggests that they adapt their interactions to the needs of their language-learning infants (Gogate, Bahrick, & Watson, 2000).

Given mothers' use of, and infants' sensitivity to, intersensory redundancy, we hypothesized that the acoustical exaggeration in mothers' ID speech – elevated pitch and expanded pitch range – would be accompanied by exaggerations in the visual prosody of speech, namely increased head movement. Furthermore, we hypothesized that links between acoustical and visual prosody would be evident in the correlations between the temporal pattern of head movements and changes in voice pitch.

Method

The procedure was designed to obtain simultaneous measures of head movements and speech acoustics while mothers talked to their infants (ID speech condition) and to an adult (AD speech condition).

Participants

Ten English-speaking mothers, with their infants, participated in this study. Mothers ranged in age from 29 to 37 years with a mean age of 31.4 years. Infants' mean age was 8.0 months ($SD = 2.3$ months). Four additional mother-infant pairs were tested, but were excluded due to infant fussiness (1), poor head-tracking quality (1), and other technical problems (2). All participants were native speakers of standard American English and had no reported speech or language deficits.

Apparatus and Procedure

Recordings of mothers' head movements were made using the head-tracking functionality of the faceLAB 4 eye-tracking system (Seeing Machines Limited, Canberra, Australia). This video-based system consists of two infrared cameras and an infrared LED light source. The faceLAB software calculated the mothers' head position by identifying and analyzing reference points (the corners of the eye and mouth) and textural features on the face. The system provides measures of head translation in three dimensions as well as measures of head rotation on three axes at a temporal resolution of 60 samples per second. The typical head tracking accuracy of the faceLAB system is ± 1 mm of translational error, and $\pm 1^\circ$ of rotational error. Simultaneous audio recordings of the mothers' speech were obtained with a lapel microphone worn by the mother and connected to a Mac Pro computer via a Roland Edirol FA-101 sound interface. Sound was recorded in AIFF format (16 bit, 44.1 kHz).

The testing was conducted in a single-wall, 10×10 foot sound booth (Industrial Acoustics Company, Inc., Bronx, NY). As illustrated in Figure 1, mothers were seated facing the infant or adult listener at a distance of between 130 to 160 cm (the distance varied over time as function of changes in mothers' head movements and posture). The faceLAB cameras were placed on small table facing the mother at a distance of about 50 cm. In the AD speech condition, the adult listeners were also seated, and were therefore at eye level with the mothers. In the ID condition, infants were seated in a high chair slightly below the mothers' eye level. This arrangement was adopted to provide a more natural physical scenario, typical of ID and AD conversational situations; elevating infants to the talkers' eye level would have forced mothers to adopt a less natural head posture when talking to their infants. The

physical presence of the infant is important for eliciting ID speech. For example, Fernald and Simon (1984) found that simulated ID speech, in which the infant was not present, did not show the same exaggerated intonation contours as ID speech.

After a short (5 minute) calibration procedure in which the faceLAB system obtained snapshots of each mother's head in several positions, in order to generate an individualized model of the mother's facial geometry, acoustical and head movement data were collected from the mother in two different conditions: ID and AD speech. Mothers were asked to spontaneously interact with their infants for 5 minutes. To avoid obstructing the camera's view of the mothers' face, no toys or objects were involved. For the AD speech condition an adult experimenter was seated across from the mother, and engaged the mother in an open-ended question about what they did on their last vacation. The mother was encouraged to expand upon her answers and prompted for additional detail until 5 minutes of recording was obtained.

Results

The data collected consisted of measures of the mothers' head translation in three dimensions and head rotation on three axes, recorded at a sampling rate of 60 Hz. Thirty-second-long excerpts were selected from the middle of the recordings in the ID and AD speech conditions. These were chosen so as to exclude periods of missing data caused by occasional movement of the head outside the tracking range or interruptions in tracking due to occlusion of the mother's face by her hand. Head movement data were low-pass filtered (cutoff = 3 Hz, reject band > 5 Hz) to remove nonbiological aspects of the signal. An example of data from one subject is shown in Figure 2. To confirm that these samples did not differ from the recordings as a whole, additional 30-second-long comparison samples (with occasional tracking interruptions) were also selected from the early and late portions of the recordings. No statistically significant differences in head movement measures were found between the comparison samples and the selected sample, which provides reassurance that the selected samples were representative of the recording as a whole. Furthermore, the absence of significant differences between the early and late comparison samples suggests that mothers' summary measures of head movements did not vary over the course of recording (e.g., due to acclimatization to the procedure), but rather varied as a function of the target listener, as will be discussed below.

Acoustical Analysis

To confirm that mothers' ID and AD speech samples were typical of those in previous studies, audio recordings for each mother in each condition were analyzed in Praat (Boersma & Weenick, 2010), using a cross-correlation method, to obtain the changing pitch (F_0) values over time. The bottom panel of the Figure 2 shows an example of pitch data for one mother. As shown in Figure 3, the mean pitch was significantly higher in the ID speech condition ($M = 268$ Hz) than in the AD speech condition ($M = 190$ Hz), $t(9) = 15.13$, $p < .0001$. The standard deviation for each mother's pitch values over time was also calculated. As expected, the mean standard deviation values across mothers was significantly higher in the ID speech condition ($M_{SD} = 75.2$ Hz) than in the AD speech condition ($M_{SD} = 57.8$ Hz),

$t(9) = 3.34, p = .009$, indicating that mothers' pitch varied more when talking to their infants.

Analysis of Head Movement

The three head-translation and three head-rotation measures provide a complete description of the changing head position and orientation over time.¹ To examine differences in the amount of head movement across conditions, the root mean square (RMS) of head translation and rotation values were calculated for each dimension of translation and each axis of rotation. These measures, shown in Figure 4, summarize the variance in mothers' head position over time, relative to their average head position over the duration of the recording, with higher RMS values indicating greater movement.

Head translation and head rotation data were analyzed separately, primarily because head translation and rotation are in different units of measurement (i.e., cm and degrees). Head translation data were examined using a two-way analysis of variance (ANOVA) with within-subjects factors of condition (ID or AD speech) and spatial dimension of head translation (X, Y or Z). A significant main effect of condition was found, $F(1,9) = 9.29$, $MSE = 0.26$, $p = 0.014$, with more overall head translation in the ID speech condition ($M = 1.32$ cm) than in the AD speech condition ($M = 0.93$ cm). This effect supports the hypothesis of exaggerated visual prosody in ID speech. A significant main effect of spatial dimension of translation was found, $F(2,18) = 19.29$, $MSE = 0.11$, $p < 0.001$, with more movement forward/back on dimension Z ($M = 1.40$ cm) than left/right on dimension X ($M = 1.20$ cm) and up/down on dimension Y ($M = 0.78$ cm). Pairwise comparisons (t -tests, with Bonferroni correction) revealed significantly greater movement in dimensions X ($p = 0.051$) and dimension Z ($p < .001$) than in dimension Y. Dimension X and Z did not differ significantly from each other ($p > .05$). No significant condition $\times$ dimension interaction was found, $F(2,18) = 1.89$, $MSE = 0.72$, $p > .05$. This suggests that although mothers showed increases in head translation when engaged in ID speech, these increases were similar across all three spatial dimensions (all $p < .05$).

Head rotation data were analyzed in the same way, using a two-way ANOVA with within-subjects factors of condition (ID or AD speech) and axis of head rotation (X, Y, Z). No significant main effect of condition was found, $F(1,9) = 1.12$, $MSE = 6.52$, $p > .05$, though there was a trend toward greater overall head rotation in the ID speech ($M = 5.42^\circ$) compared to the AD speech condition ($M = 4.72^\circ$). A main effect of axis of rotation was found, $F(2,18) = 3.57$, $MSE = 4.65$, $p = 0.049$, with more movement on the X ($M = 5.78^\circ$) and Y ($M = 5.39^\circ$) axes, than the Z axis ($M = 4.04^\circ$), though the associated pairwise comparisons were not statistically significant. No significant condition $\times$ axis interaction was found, $F(2,18) = 3.38$, $MSE = 2.80$, $p = 0.057$.

¹Although decomposing natural head movements into various orthogonal components provides a useful analytical method, it is important for a number of reasons not to assume that the analytical framework mirrors the way in which mothers generate and control their head movements. First, movement of the head on a single axis or dimension is not possible. For example, because the head sits at the top of a spine, tilting the head to the right also translates the volume of space surrounding the head to right. Second, the head position is measured relative to a spatial origin between the two recording cameras directed towards the mother, between her and her infant (see Figure 1). Although this camera position was chosen to optimize tracking quality, it is in a sense arbitrary relative to the dyad. As a result, the same head translation forward can be represented on one dimension if the movement is directly toward the origin, or two dimensions if the origin is slightly off to the side.

Overall, the head translation and rotation data show increased head movement, or exaggerated visual prosody, for ID speech. Although the interaction effects of speech condition and dimension/axis of movement were not significant, the increased forward/back translation (on dimension Z) and up/down rotation (on axis X) in the ID speech conditions suggests increased movement in the sagittal plane consistent with more prevalent “approach” movements and nods. Similarly, the increased left/right translation and rotation (on dimension X and axis Y) correspond to greater propensity to head shaking.

Relations between pitch (F_0) and head movement over time

The next analysis examined how changes in the acoustical pitch (F_0) over time relate to changes in head position over time. The goal of this analysis was to determine whether higher or lower values on certain head movement measures might coincide in time with points of higher F_0 . The F_0 data derived from mothers’ speech were time aligned with the related head translation and rotation data, using a resampling procedure in Matlab. This was necessary because the head movement data consisted of 60 data points per second, and the acoustical pitch analysis estimated F_0 , 100 times per second. The resampling procedure essentially decreased the sampling rate of the pitch data to match the head movement data, by interpolating between data points as necessary.

Because the F_0 data contained several intervals during which pitch could not be estimated (e.g., during breaths, silences, stop consonants), a subset of time points at which both head movement data and F_0 data were available were extracted. The number of common time points varied across subjects, with an average 802 data points for AD speech segments and 686 data points for ID speech. These time points represent 45% and 38% of possible time points for the AD and ID speech conditions, respectively.

These data, for each individual subject and each condition, were submitted to a multiple regression analysis to determine the degree to which F_0 could be predicted by the 6 head movement variables (3 axes of rotation and 3 dimensions of translation). The primary measure of interest here was the R^2 value which describes the proportion of variance in F_0 collectively accounted for by the six head movements variables. The regression analyses were performed in time windows that varied in size. The largest, 30-second, window included the entire sample and produced a single R^2 value. For the 15-second window, separate analyses were performed on the first and second 15-second-long halves of the sample, and the two resulting R^2 values were averaged for each subject in each condition. For 10-second-long windows, the sample was divided into thirds, and so forth for the remaining 6×5 -sec, 10×3 -sec, 15×2 -sec, and 30×1 -sec windows. As the windows decreased in size, the number of samples containing pitch and head movement data also decreased. Windows containing fewer than 10 samples were excluded from the analysis, and the R^2 values were averaged across the windows that were not excluded.

This analysis produced a single R^2 value for each subject in each condition for each time window length. These R^2 values were then submitted to a two-way repeated measures ANOVA with speech condition and time window as factors. Mean R^2 values for the ID and AD speech conditions, over the range of time windows are shown in Figure 5. A significant main effect of speech condition was found, $F(1,9) = 9.303$, $MSE = .069$, $p = .0138$, with

head movement accounting for 55.9% of the variance in pitch in the ID condition, and 42.4% in the AD speech condition. A significant main effect of time window was found, $F(1,9) = 9.303$, $MSE = .069$, $p = .0138$, with higher R^2 values for smaller time windows. The interaction of condition $\times$ time window was not significant, $F(1,9) = 0.257$, $MSE = .017$, *ns*. These results demonstrate that head movement and voice pitch are more closely related in ID speech than in AD speech, and that these relations are stronger at short time scales.

Next, the relation between each individual head movement variable and pitch were examined using simple correlations, using the same windowing procedure. For each subject and for each condition, correlation coefficients were calculated between the time-varying F_0 values and head translation and rotation measures for the 30-second sample as a whole, as well as the average coefficient across smaller windows. Figure 6 shows two versions of the average correlation between pitch and head movement as a function of time window and speech condition (ID or AD speech): the average correlation, r , and the average *absolute* correlation, $abs(r)$, in which the direction of the relation (positive or negative) is ignored. For short time windows of a few seconds absolute correlations of around, $abs(r) = .40$, indicate the presence of moderate local relations between individual head movement variables and F_0 . However, the decline in strength of these correlations as the window increases in length suggests that the particular form of the relation between head movement and pitch is less consistent across longer time intervals. This variability in the form of the relation is also evident when the average $abs(r)$ values are compared with the average r values in which the directionality of the correlation is taken into account. Across all time windows the average correlation hovers around $r = .00$, and this difference relative to the absolute correlation reflects the cancelling out of correlations that are sometimes positive and sometimes negative when they are averaged.

Discussion

The present study provides a quantitative examination of the head movements underlying ID visual prosody and its relation to the well-known acoustical prosody of ID speech. Mothers' exaggerated acoustical prosody, characterized by an expanded pitch range, is to a degree paralleled visually, with increased head movement. This demonstration of ID visual prosody, along with other forms of ID facial movements and gestures (Brand et al., 2002; Brand & Shallcross, 2008; Chong, Werker, Russell, & Carroll, 2003; Cohn & Tronick, 1988; Messinger, Mahoor, Chow, & Cohn, 2009), provides additional evidence that mothers adapt their behavior, using a broad repertoire, when interacting with infants.

Head movements likely reflect a number of levels of communicative function. At one level they too may be related to the physical production of speech, for which one might expect strong coupling of movements with speech acoustics. In contrast, head movements may serve other expressive functions, less related to speech articulation and acoustics. In the present study, regression analyses found that large proportions of the variance in voice pitch over time can be accounted for by visual prosodic head movements. However, these relations are more local in nature, being strongest over shorter time windows, and the correlations between head movement variables and F_0 , though moderate in strength in absolute terms, vary in polarity across time windows.

Under different conditions, other studies have found strong relations between head movements and speech acoustics. For example, Munhall et al. (2004) found that over 63% of the variance in F_0 could be accounted for by head movement measures, though comparison between these studies is difficult given numerous experimental differences. The present study examines a variety of mothers producing spontaneous speech in the context of dyadic interaction, whereas the head movement data in Munhall et al.'s study were obtained from a male native speaker of Japanese reciting predefined sentences. A related earlier study by Yehia, Kuratate and Vatikiotis-Bateson (2002) found strong correlations between F_0 and head movement within individual utterances, but the way in which these measures were coupled varied from utterance to utterance, and even between repetitions of the same utterance by the same talker, causing the correlations to disappear entirely when calculated for the corpus as a whole. This reduced corpus-wide correlation in Yehia et al.'s study corresponds to our observed decrease in R^2 values for larger time windows, as well as the decreased average correlations, r , in comparison to the average absolute correlations, $abs(r)$. Altogether, the present and previous work demonstrates relations between acoustical and visual prosody in speech, but that these relation are more local in time span, with less consistency over longer periods.

In the present study, the residual variance in F_0 not related to head movements suggests that mothers' expressive head movements are not a simple byproduct of speech production. These head movements may serve other functions. One possibility is that ID visual prosody may facilitate speech perception in infants, in the same way that extraoral visual information increases the intelligibility of speech in adult listeners (Davis & Kim, 2006; Munhall et al., 2004; Thomas & Jordan, 2004). This enhancement may rely on correlations between head movement and stress patterns in speech (Hadar et al., 1983; Hadar, Steiner, Grant, et al., 1984). Although, infants are sensitive to audiovisual correspondences in speech (Kuhl & Meltzoff, 1982; Patterson & Werker, 2002, 2003) and have shown enhancement in learning with visual speech (Teinonen et al., 2008), it remains unclear the degree to which this enhancement relies on oral versus extraoral visual information (Lewkowicz & Hansen-Tift, 2012; Smith, Gibilisco, Meisinger, & Hankey, 2013). The finding that infants are able to match the emotional expression of talking faces with emotional speech, even when the area around the mouth is obscured (Walker-Andrews, 1986), suggests that infants use extraoral visual information in speech perception.

More generally the prosodic head movements may serve to emphasize or attract and facilitate infants' attention to a particular stimulus, event or word. Understanding the particular role that visual prosodic head movements play in this function is difficult because in natural contexts, head movements are so closely integrated with facial movements related to speech articulation, emotional expression, mothers' eye gaze, as well and speech acoustics and linguistic meaning.

This work could be extended in a number of ways to better understand the function of these extraoral head movements. One is to focus on maternal production by better controlling mothers' communicative goals to see how mothers might use head movements to highlight particular words, linguistic constructs or pragmatic functions in their speech. A complementary approach would be to focus on the infants' perception and understanding of

these expressive head movements to determine how they might affect attention, emotional responses and learning. Teasing apart the role of head movements, independent of other facial movements and expression is difficult. However, one potential approach might involve creating stimuli consisting of facial animations whose movements follow the movement patterns of real mothers' ID visual prosody to examine infants' responses to ID visual prosody in a context in which other factors can be controlled. Overall, the present study provides evidence for an additional component to mothers' already broad repertoire of expressive behaviors and adaptations employed when interacting with their young children.

Acknowledgments

We thank Colleen Gibilisco for assistance with subject recruitment and data collection. This research was supported in part by the NIH (T35DC008757; P30DC4662).

References

- Bahrack LE, Pickens JN. Classification of bimodal English and Spanish language passages by infants. *Infant Behavior and Development*. 1988; 11:277–296.
- Boersma, P.; Weenick, D. Praat: doing phonetics by computer (Version 5.2) [Computer Program]. 2010.
- Brand RJ, Baldwin DA, Ashburn LA. Evidence for 'motionese': modifications in mothers' infant-directed action. *Developmental Science*. 2002; 5:72–83.
- Brand RJ, Shallcross WL. Infants prefer motionese to adult-directed action. *Developmental Science*. 2008; 11:853–861. [PubMed: 19046154]
- Chong S, Werker J, Russell &, Carroll J. Three facial expressions mothers direct to their infants. *Infant and Child Development*. 2003; 12:211–232.
- Cohn J, Tronick E. Mother-infant face-to-face interaction: Influence is bidirectional and unrelated to periodic cycles in either partner's behavior. *Developmental Psychology*. 1988; 24:386–392.
- Cooper RP, Aslin RN. Preference for infant-directed speech in the first month after birth. *Child Development*. 1990; 61:1584–1595. [PubMed: 2245748]
- Davis C, Kim J. Audio-visual speech perception off the top of the head. *Cognition*. 2006; 100:21–31.
- Fernald A. Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*. 1985; 8:181–195.
- Fernald A, Kuhl P. Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*. 1987; 10:279–293.
- Fernald A, Simon T. Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*. 1984; 20:104–113.
- Fernald A, Taeschner T, Dunn J, Papousek M, de Boysson-Bardies B, Fukui I. A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*. 1989; 16:477–501. [PubMed: 2808569]
- Gogate LJ, Bahrack LE, Watson JD. A study of multimodal motherese: the role of temporal synchrony between verbal labels and gestures. *Child Development*. 2000; 71:878–894. [PubMed: 11016554]
- Gogate LJ, Walker-Andrews A. More on developmental dynamics in lexical learning. *Developmental Science*. 2001; 4:31–37.
- Gogate LJ, Walker-Andrews A, Bahrack LE. The intersensory origins of word comprehension: an ecological-dynamic systems view. *Developmental Science*. 2001; 4:1–37.
- Graf, HP.; Cosatto, E.; Strom, V.; Huang, FJ. Visual prosody: Facial movements accompanying speech. Paper presented at the Fifth IEEE International Conference on Automatic Face and Gesture Recognition (FGR '02); Washington, D.C. 2002.
- Green JR, Nip ISB, Wilson EM, Mefferd AS, Yunusova Y. Lip movement exaggerations during infant-directed speech. *Journal of Speech, Language, and Hearing Research*. 2010; 53:1529–1542.

- Hadar U, Steiner TJ, Grant EC, Rose FC. Head movement correlates of juncture and stress at sentence level. *Language and Speech*. 1983; 26:117. [PubMed: 6664179]
- Hadar U, Steiner TJ, Grant EC, Rose FC. The timing of shifts of head posture during conversation. *Human Movement Science*. 1984; 3:237–245.
- Hadar U, Steiner TJ, Rose FC. Involvement of head movement in speech production and its implications for language pathology. *Advances in neurology*. 1984; 42:247. [PubMed: 6209950]
- Hollich G, Newman RS, Jusczyk P. Infants' use of synchronized visual information to separate streams of speech. *Child Development*. 2005; 76:598–613. [PubMed: 15892781]
- Hollich G, Prince CG. Comparing infants' preference for correlated audiovisual speech with signal-level computational models. *Developmental Science*. 2009; 12:379–387. [PubMed: 19371360]
- Kamachi M, Hill H, Lander K, Vatikiotis-Bateson E. 'Putting the face to the voice': Matching identity across modality. *Current Biology*. 2003; 13:1709–1714. [PubMed: 14521837]
- Kuhl PK, Meltzoff A. The intermodal representation of speech in infants. *Infant Behavior and Development*. 1984; 7:361–381.
- Kuhl PK, Meltzoff AN. The bimodal perception of speech in infancy. *Science*. 1982; 218:1138–1141. [PubMed: 7146899]
- Lewkowicz DJ, Hansen-Tift AM. Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceeding of the National Academy of Sciences*. 2012; 109:1431–1436.
- Masataka N. Perception of motherese in Japanese sign language by 6-month-old hearing infants. *Developmental Psychology*. 1998; 34:241–246. [PubMed: 9541776]
- Messinger D, Mahoor MH, Chow S, Cohn JF. Automated measurement of facial express in infant-mother interaction: a pilot study. *Infancy*. 2009; 14:285–305. [PubMed: 19885384]
- Munhall K, Jones JA, Callan DE, Kuratate T, Vatikiotis-Bateson E. Visual prosody and speech intelligibility: head movement improves auditory speech perception. *Psychological Science*. 2004; 15:133–137. [PubMed: 14738521]
- O'Neill M, Bard K, Linnell M, Fluck M. Maternal gestures with 20-month-old infants in two contexts. *Developmental Science*. 2005; 8:352–359. [PubMed: 15985069]
- Patterson ML, Werker JF. Infants' ability to match dynamic phonetic and gender information in the face and voice. *Journal of Experimental Child Psychology*. 2002; 81:93–115. [PubMed: 11741376]
- Patterson ML, Werker JF. Two-month-old infants match phonetic information in lips and voice. *Developmental Science*. 2003; 6:191–196.
- Pegg J, Werker J, McLeod P. Preference for infant-directed over adult-directed speech: Evidence from 7-week-old infants. *Infant Behavior and Development*. 1992; 15:325–325.
- Rosenblum LD, Schmuckler MA, Johnson JA. The McGurk effect in infants. *Perception & Psychophysics*. 1997; 59:347–357. [PubMed: 9136265]
- Smith L, Gasser M. The development of embodied cognition: Six lessons from babies. *Artificial Life*. 2005; 11:13–29. [PubMed: 15811218]
- Smith NA, Gibilisco CR, Meisinger RE, Hankey M. Asymmetry in infants' selective attention to facial features during visual processing of infant-directed speech. *Frontiers in Psychology*. 2013; 4:601. [PubMed: 24062705]
- Smith NA, Trainor LJ. Infant-directed speech Is modulated by infant feedback. *Infancy*. 2008; 13:410–420.
- Stern D, Spieker S, Barnett R, MacKain K. The prosody of maternal speech: infant age and context related changes. *Journal of Child Language*. 1983; 10:1. [PubMed: 6841483]
- Stern D, Spieker S, MacKain K. Intonation contours as signals in maternal speech to prelinguistic infants. *Developmental Psychology*. 1982; 18:727–735.
- Teinonen T, Aslin R, Alku P, Csibra G. Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*. 2008; 108:850–855. [PubMed: 18590910]
- Thomas SM, Jordan TR. Contributions of Oral and Extraoral Facial Movement to Visual and Audiovisual Speech Perception. *Journal of Experimental Psychology: Human Perception and Performance*. 2004; 30:873–888. [PubMed: 15462626]

- Thompson WF, Russo FA. Facing the music. *Psychological Science*. 2007; 18:756–757. [PubMed: 17760767]
- Walker-Andrews A. Intermodal Perception of Expressive Behaviors: Relation of Eye and Voice? *Developmental Psychology*. 1986; 22:373–377.
- Weikum WM, Vouloumanos A, Navarra J, Soto-Faraco S, Sebastián-Gallés N, Werker JF. Visual language discrimination in infancy. *Science*. 2007; 316:1159. [PubMed: 17525331]
- Werker JF, Pegg J, McLeod P. A cross-language investigation of infant preference for infant-directed communication. *Infant Behavior and Development*. 1994; 17:323–333.
- Yehia H, Kuratate T, Vatikiotis-Bateson E. Linking facial animation, head motion and speech acoustics. *Journal of Phonetics*. 2002; 30:555–568.
- Yehia H, Rubin P, Vatikiotis-Bateson E. Quantitative association of vocal-tract and facial behavior. *Speech Communication*. 1998; 26:23–43.

Figure 1.

An illustration of the experimental setup in the infant-directed speech condition. The eye-tracker cameras on the table recorded the mother's head movements as she spoke to her infant.

Figure 2.

Head movement and voice pitch data from a representative individual mother during a 30-second-long sample of infant-directed speech. The top six panels show changes in the mother's head position on three dimensions of translation and three axes of rotation. The bottom panel shows changes in the mother's voice pitch (F_0).

Figure 3.

Pitch measures (F_0) for mothers' infant-directed (ID) and adult-directed (AD) speech. The left panel shows the mean of individual mothers' mean pitch values. The right panel shows the mean of the standard deviation in pitch values across mothers. Error bars reflect the standard error of the mean.

Figure 4. RMS (root-mean-square) measures of head translation on three dimensions (X, Y, Z), and head rotation on three axes (X, Y, Z) for mothers' infant-directed (ID) and adult-directed (AD) speech. Error bars show the standard error of the mean.

Figure 5.

Results of the multiple regression analysis. The average R^2 values show the proportion of the variance in voice pitch accounted for by the six head movement measures in the infant-directed (ID) and adult-directed (AD) speech conditions. For the 30-second time window, the value reflects the single R^2 value for the sample as a whole. For the shorter time windows, the value represents the average R^2 across multiple smaller windows. Standard error bars are shown.

Figure 6.

Correlations between the six head position variables and changing voice pitch calculated over a range of time windows. Each filled point represents the correlation coefficient, r , averaged across mothers in the infant-directed (ID) and adult-directed (AD) speech conditions. The hollow points represent the average *absolute* correlation coefficient, $abs(r)$, in which the direction of the correlation (positive or negative) is ignored. Standard error bars are shown.