HHS Public Access

Author manuscript

Dev Sci. Author manuscript; available in PMC 2024 September 01.

Published in final edited form as:

Dev Sci. 2023 September ; 26(5): e13359. doi:10.1111/desc.13359.

Infant-Directed Song Potentiates Infants' Selective Attention to Adults' Mouths over the First Year of Life

Camila Alviar[1], **Manash Sahoo**[2,3], **Laura Edwards**[2,3], **Warren Jones**[2,3], **Ami Klin**[2,3], **Miriam Lense**[1,4,5]

[1]Department of Otolaryngology - Head & Neck Surgery, Vanderbilt University Medical Center, Nashville, TN, USA

[2]Marcus Autism Center, Children's Healthcare of Atlanta, Atlanta, GA, USA

[3]Emory University School of Medicine, Atlanta, GA, USA

[4]Vanderbilt Kennedy Center, Vanderbilt University Medical Center, Nashville, TN, USA

[5]The Curb Center for Art, Enterprise, and Public Policy, Vanderbilt University, Nashville, TN, USA.

Abstract

The mechanisms by which infant-directed speech and song support language development in infancy are poorly understood, with most prior investigations focused on the auditory components of these signals. However, the visual components of infant-directed communication are also of fundamental importance for language learning: over the first year of life, infants' visual attention to caregivers' faces during infant-directed *speech* switches from a focus on the eyes to a focus on the mouth, which provides synchronous visual cues that support speech and language development. Caregivers' facial displays during infant-directed *song* are highly effective for sustaining infants' attention. Here we investigate if infant-directed song specifically enhances infants' attention to caregivers' mouths. 299 typically developing infants watched clips of female actors engaging them with infant-directed song and speech longitudinally at six time points from 3–12 months of age while eye-tracking data was collected. Infants' mouth-looking significantly increased over the first year of life with a significantly greater increase during infant-directed song versus speech. This difference was early-emerging (evident in the first 6 months of age) and sustained over the first year. Follow-up analyses indicated specific properties inherent to infant-directed song (e.g., slower tempo, reduced rhythmic variability) in part contribute to infants' increased mouth-looking, with effects increasing with age. The exaggerated and expressive facial

Correspondence regarding this article should be addressed to Camila Alviar (maria.c.alviar-guzman@vumc.org) or Miriam Lense (miriam.lense@vanderbilt.edu). Music Cognition Lab, 1408 17th Ave. S, Nashville, TN 37212.

Conflict of interest disclosure

WJ and AK are scientific consultants to and minority shareholders in EarliTec Diagnostics, Inc. EarliTec develops technology for early identification of autism and gives revenue to support treatment of children with autism. The activity has been reviewed and approved by Emory University's Conflict of Interest Review Office. WJ and AK's work with EarliTec is unrelated to the present paper.

Ethics approval statement

The research protocol was approved by Human Investigations Committees at Yale and Emory University Schools of Medicine, as well as Children's Healthcare of Atlanta. All parents and/or legal guardians gave written informed consent.

features that naturally accompany infant-directed song may make it a particularly effective context for modulating infants' visual attention and supporting speech and language development in both typically developing infants and those with or at risk for communication challenges.

Keywords

infant-directed speech; infant-directed song; language development; infants; eye-tracking

Introduction

Infant-directed (ID) communication frequently occurs face-to-face allowing infants to both see and hear their caregivers as they engage with them. This multimodality offers infants meaningful and redundant cues that support social interaction and language learning (Bahrick et al., 2019; Flom & Bahrick, 2007). Cues to parse language structure, and understand speakers' intentions and affective states, are found not only in the exaggerated acoustics and prosody of caregivers' vocalizations (Bryant & Barret, 2007; Falk & Audibert, 2021; Fernald, 1989; Papousek et al., 1991), but also in their corresponding facial expressions and movements. Facial movements carry echoes of the configurations of the vocal tract (Yehia et al., 2018); lip aperture and jaw displacement closely match the acoustic envelope of the caregivers' vocalizations (Chandrasekaran et al., 2009), and also convey information about the affective states of the speaker (Livingstone et al., 2015; Tartter, 1980). Caregivers' exaggerated facial expressions additionally provide communicative information (Chong et al., 2003; Shepard et al., 2012). Eyebrow movements and head movements accompany and highlight prosodic phrase boundaries (de la Cruz-Pavía et al., 2020; Swerts & Krahmer, 2008) and portray the speaker's emotional intent (Livingstone et al., 2015; Livingstone & Palmer, 2016). The eyes of the caregiver offer information on the caregiver's affect while their gaze direction is key to establishing joint attention (Brooks & Meltzoff, 2002, 2005).

During early childhood, infants' attention to different elements of this rich array of visual cues changes across their developmental trajectory. When engaged by audiovisual displays of ID speech, for example, infants preferentially look at the eyes of the speaker during the first months of life, and slowly shift their attention to the mouth during the second half of their first year, a developmental period associated with growth in the infant's own communicative skills and emerging linguistic repertoire (e.g., start of babbling; Jones & Klin, 2013; Lewkowicz & Hansen-Tift, 2012; Tenenbaum et al., 2013; Wagner et al., 2013). This increased interest in the mouth during the second half of the first year of life likely takes advantage of redundant and synchronized audio and visual cues that support language learning (Hillairet de Boisferon et al., 2017; Lewkowicz & Hansen-Tift, 2012; Tenenbaum et al., 2013). Infants' attention to the mouth region of an engaging speaker at 6 and 12 months of age predicts their concurrent expressive language development (Tsang et al., 2018), as well as later expressive language development at 18 and 24 months (Tenenbaum et al., 2015; Young et al., 2009). Similarly, attention to the mouth during ID speech at 6 months predicts later receptive language development at 12 months (Imafuku & Myowa, 2016).

Like ID speech, infants' experiences with ID song are associated with their language and communication development including gesture use (Gerry et al., 2012; Papadimitriou et al., 2021), receptive language (Papadimitriou et al., 2021), and vocabulary (Franco et al., 2021). However, infants' visual attention allocation during ID song is less studied, despite song being ubiquitous in infants' communicative environments (Steinberg et al., 2021; Trehub et al., 1997, Yan et al., 2021). Compared to adult-directed speech, ID speech already involves many characteristics that make it more musical and song-like, such as slower tempo, increased repetitiveness and rhythmicity, and exaggerated and more positive pitch contours and facial expressions and head movements (Chong et al., 2003; Fernald et al., 1989; Grieser & Kuhl, 1988; Stern, 1974; Stern et al., 1983). All these features attract and maintain infants' overall attention (and attention to the speaker's mouth) more during ID speech than adult-directed speech (Fernald, 1985; Lewkowicz & Hansen-Tift, 2012; Werker & McLeod, 1989; Werker et al., 1994). Compared to ID speech, however, ID song captures infants' attention faster and sustains it for longer durations during multimodal presentations (via live interactions or audio-video recordings) or visual-only presentations (Macari et al., 2021; Trehub et al., 2016; although see Costa-Giomi, 2014) but not audio-only presentations (Corbeil et al., 2013; Costa-Giomi, 2014; Costa-Giomi & Ilari, 2014), suggesting that visual features play an important role in modulating infants' engagement with the communicative signal.

ID song expands upon many of the features of ID speech that are believed to be important for infant attention regulation. ID song is slower, more rhythmic (Hilton et al., 2022; Trainor, 1996), more routinized across contexts and interactions (Bergeson & Trehub., 2002; Kragness et al., 2022; Mendoza & Fausey, 2021), and therefore more predictable, than ID speech. Song in general features larger jaw movements accompanied by increased amplitude compared with speech (Livingstone et al., 2015); in conjunction with the slower and more rhythmic qualities of song (Ding et al., 2017), this suggests that audiovisual synchrony in the mouth area is more pronounced for ID song than ID speech. ID song also involves positive affect more consistently and frequently than ID speech: in Western cultures (the focus of the current study), mothers sing playful songs to their infants more often than lullabies (Trehub et al., 1997), and smile more while singing than while speaking to their infants (Trehub et al., 2016). In addition to these attention-regulating attributes, ID song is highly effective at modulating infants' arousal levels (Corbeil et al., 2016; Nakata & Trehub, 2004; Trehub et al., 2015; Tsang et al., 2017). Infants calm faster and for longer periods of time in response to ID song than ID speech, particularly for familiar songs and positive, playful songs (Cirelli & Trehub, 2020; Corbeil et al., 2016; Trehub et al., 2015).

The differing attributes and contextual effects of ID song and ID speech imply two possible but competing hypotheses with regards to infants' attention allocation to facial visual cues during song as compared to speech. On one hand, several features of ID song might increase attention to the mouth: infants show early sensitivity to amodal properties such as tempo, rhythm, synchrony, and affect. These properties are highly salient in a caregiver's mouth region during speaking and singing due to the tight links between orofacial movements and vocal production (Bahrick et al., 2004; Flom & Bahrick, 2007; Lewkowicz, 2003; Lewkowicz & Marcovitch, 2006) and these features tend to be enhanced during song versus speech (Livingstone et al., 2015; Trainor, 1996; Trehub et al., 2016; Tsang et al., 2017).

Prior work with experimentally-manipulated child-directed speech indicates that slower tempo speech increases duration of fixations to the mouth in young children (Gepner et al., 2021) while audiovisual synchrony mediates mouth-looking during late infancy (Hillareit de Boisferon et al., 2017, though note that both these effects may in part be the result of odd or surprising stimuli). On the other hand, the special role of song in emotion regulation and social bonding (Cirelli et al., 2020; Corbeil et al., 2016; Trainor, 1996), and the role of the eyes in communicating social and deictic information (Brooks & Meltzoff, 2002, 2005; Buchan et al., 2007; Symmons et al., 1998; Tomasello et al., 2007), would predict comparable, if not reduced, amounts of mouth-looking—in favor of eye-looking—in ID song versus ID speech.

Comparing infants' facial scanning behaviors during ID song and ID speech informs how different communicative contexts impact infants' visual attention beyond the overall attentional capture effects of these salient and meaningful interactions. The different properties of ID song and ID speech create a natural opportunity to investigate how specific communicative features may underlie infant facial attention allocation, which may elucidate the mechanisms by which these two communicative contexts support language and communication development. In the current study, we conducted a secondary data analysis of an extant longitudinal dataset of visual attention in infants from 3 to 12 months, to address the possibility of different facial scanning patterns to ID song and ID speech over the first year of life. Specifically, we compared infants' allocation of visual attention to an actor's mouth when being engaged by ID song and ID speech over the first year of life. We focused on infant's mouth-looking based on prior findings regarding changes in mouth-looking to ID speech over this time period, and on theorized relationships between mouth-looking and language development (de Boisferon et al., 2017; Lewkowicz & Hansen-Tift, 2012; Tenenbaum et al., 2013). We additionally quantified and tested the influence of features inherent to the communicative signal but that vary across song and speech, and which are particularly observable from an interlocutor's mouth movements—tempo, rhythmicity, audiovisual synchrony, and positive affect—on infants' preference for the mouth over the first year of life across the ID song and ID speech contexts.

Methods

Participants

We reanalyzed an existing dataset of 299 typically developing infants (155 male, 144 female) who were eye-tracked longitudinally at 3, 4, 5, 6, 9, and 12 months of age as part of a larger study on social development. All children with usable eye-tracking data were included in the study regardless of total number of usable visits, however 73% of the infants had usable data from at least 3 visits ($M = 3.51$). More details about the sample can be found in the Supplementary Methods. Only typically developing infants with no concerns for developmental or intellectual disabilities or familial history of autism spectrum disorder (ASD) in first, second, or third-degree relatives were included in the current study. All participants were screened for normal or corrected-to-normal vision and for normal hearing using medical and developmental history, otoacoustic emissions testing for hearing, and basic tests of visual function including the ability to shift and stabilize gaze. Participants

were recruited from general OB/GYN and primary pediatric care practices in the community and parent social networks. The research protocol was approved by Human Investigations Committees at Yale and Emory University Schools of Medicine, as well as Children's Healthcare of Atlanta, and all parents and/or legal guardians gave written informed consent.

Stimuli

The total stimuli set consisted of up to 16 possible audiovisual clips showing one of five female actors looking directly into the camera and engaging the child with either ID song (6 clips) or ID speech (10 clips) against a nursery background (see Figure 1A). Clips were designed with the goal of naturalistic validity and aimed to capture a broad range of common childhood experiences: the speech clips depicted short excerpts of typical care routines (e.g., playtime, mealtime), while the song clips consisted of common infant-directed songs sung in a playful manner (e.g., "Old MacDonald", "Twinkle Twinkle Little Star"). Clip duration ranged from 9.8 to 43.4 seconds ($M = 21.3$, $SD = 6.2$). A summary of acoustic and visual features for both clip types (ID song and ID speech) is shown in Table 1 and Figure 2, and each measure is briefly described below. More technical specifications of the stimuli and their presentation can be found in the Supplementary Methods.

The specific clips included in the playlists varied across eye-tracking session age points but each session playlist included at least 3 song clips and 6 speech clips (see Supplementary Methods for information on repeating versus novel clips). Despite clip variation across playlists, the distributions of clip characteristics (see below: tempo, rhythmicity, salience of audiovisual synchrony in the mouth, and positive affect) across age points stayed relatively stable overall, as well as within contexts (i.e., speech and song; see Figure S2 and detailed analyses in the Supplementary Methods). Children provided usable data for an average of 2.4 (SD = 1.0) song clips and an average of 3.9 (SD = 2.0) speech clips in a given eye-tracking session (see Procedures below).

Clip Features—We characterized clips via a range of acoustic, visual, and audiovisual features to quantify physical attributes of ID song and speech reflected in our stimuli (Table 1, Figure 2).

Pitch.—We calculated the fundamental frequency of each clip using the MATLAB "pitch" function (Mathworks). To avoid noise and edge effects, intervals of silence were removed from the fundamental frequency time-series, and each series was smoothed with a 1ms-span median filter. Series were manually inspected to confirm there were no octave errors. The mean and standard deviation were derived for each clip from the entire fundamental frequency time-series to obtain average pitch and pitch variability across clips.

Tempo.—We calculated tempo as the number of syllables spoken or sung per second. We transcribed the words in each clip, and then used the Carnegie Mellon Pronouncing Dictionary (Weide, 1998) to automatically determine the total number of syllables in each clip. We then divided that number by the duration of the clip in seconds to obtain a measure of syllables per second.

Rhythmic Variability.—We measured the rhythmic variability in each clip using the normalized Pairwise Variability Index (nPVI) of the vowel durations. Vowel onsets and offsets were annotated by hand in Praat (Boersma & Weenink, 2022), and nPVI was calculated as the sum of the normalized duration differences between consecutive vowels. nPVI has been used previously to quantify and compare rhythm across speech and music (e.g., Hannon et al., 2016; Patel et al., 2006).

Salience of Audiovisual Synchrony (AVS) in the Mouth vs the Eyes.—For each pair of consecutive frames of a clip, we calculated the AVS in the eyes and mouth regions of interest (ROI, see Figure 1B) as the product of the optic flow in each ROI (i.e., amount and direction of change in brightness from one frame to the next) and the average root-mean-square (RMS) of the amplitude envelope corresponding to that pair of frames. We summed the AVS in each ROI across frames to obtain the total AVS per ROI for each clip. To determine whether the AVS was more salient in the eyes or the mouth in a given clip, we divided the AVS in the mouth by the AVS in the eyes to obtain a ratio. Values higher than 1 indicate higher AVS in the mouth area. Optic flow was calculated using scripts adapted from Matlab's Optic Flow toolbox (Karlsson, 2022).

Positive Affect.—We quantified positive affect in clips as the percentage of frames in which the actor was smiling. Frame-by-frame presence of smiling was determined by applying OpenFace (Baltrusaitis et al., 2018; Baltrusaitis et al., 2015) recognition algorithms to measure the presence of Facial Action Units 12 (lip corner puller) and 6 (cheek raiser), which are active when a person smiles (Schmidt & Cohn, 2001). For each frame of a clip, OpenFace assigned a 1 if a given action unit was present and a 0 if it was absent. For each clip, we counted the number of frames in which both action units were identified as present and divided them by the total number of frames in the clip.

Consistent with the goals of the larger study, and as noted above, ID song and speech clips were designed to reflect naturally occurring caregiving interactions. As illustrated in Figure 2, the song and speech clips followed expected canonical patterns: ID song clips were, on average, slower in tempo, less rhythmically variable, showed higher saliency of AVS in the mouth than the eyes area, and had higher positive affect than the speech clips (Table 1, Figure 2). Average pitch and pitch variability (F_0 mean and standard deviation) were similar across ID song and ID speech clips. These featural differences and similarities are consistent with prior studies of maternal ID song and speech and reflect the ecological validity of the current stimuli (Hilton et al., 2022; Trainor, 1996).

Procedure

A full description of all experimental procedures, technical specifications of the experimental stimuli, calibration procedures, data acquisition, and data coding protocols can be found in the Supplemental Methods. In brief, participants completed eye-tracking protocols in a dedicated testing room where they were shown familiar, engaging videos (e.g., Elmo) while becoming comfortably situated. Eye-tracking equipment was then calibrated to each infant using a 5-point calibration scheme with targets presented on an otherwise blank screen. Calibration was within 3° of target center across ages (see Figure S1 and Figure S2

in Supplemental Method). Once calibration was complete, children were presented with the audiovisual ID song and ID speech clips, interlaced with other clips not analyzed in this study. The clips were presented in the same pseudo-random order to all children within an age point. The selection and number of video clips, as well as the presentation order, varied across data collection time points, in order to maximize developmental appropriateness and participant engagement. The clips in a playlist were pseudo-randomly selected to (a) strike a balance of novelty and repetition, with 70% of the clips being repeated from previous playlists; and (b) prevent several clips of the same type from appearing consecutively (see Supplementary Methods for detailed analyses of playlist composition over time).

Analysis Plan

Fixations were coded into four ROIs: Eyes, Mouth, Body, and Object (Figure 1B). Percentage of mouth-looking for each clip was quantified as the proportion of face-looking time (PFLT) spent on the actor's mouth (PFLT-m); that is, the duration of all fixations to the mouth ROI divided by the durations of all fixations to the face (eyes + mouth ROIs). This metric forefronts the eyes-mouth trade-off in infants' visual attention and facilitates comparison with previous literature that uses the same metric (Ayneto & Sebastian-Galles, 2016; Hillaret de Boisferon et al., 2018; Imafuku & Myowa, 2016; Lewkowicz & Hansen-Tift, 2012; Pons et al., 2019; Tsang et al., 2018). Note that PFLT-m is the complement of attention to the eyes as proportion of face-looking time, and so the results can also be easily interpreted with respect to eye-looking (i.e., 60% *mouth*-looking can also be interpreted as 40% *eye*-looking).

We used a series of mixed-effects models to examine the effects of age, clip type (ID song vs. ID speech), and clip features on infants' mouth-looking (PFLT-m). In the first model we investigated the effects of age, clip type, and their interaction to test both developmental change, and differences in mouth-looking to ID song and ID speech. In the second model, we exchanged clip type with the array of clip features under study: tempo, rhythmic variability, mouth AVS saliency, and positive affect, and their interactions with age, to test possible drivers of mouth-looking differences across song and speech (we had no a priori reason to believe pitch might affect attention to the mouth on its own, so it was not included in the model predictors).

We controlled for sex in all models and included a random intercept and slope for age for each child to account for individual differences in PFLT-m in the developmental trajectory. Age was centered ($M = 6.6$ months), clip type and sex were contrast coded with simple effects coding, and all clip features were Z-scored to make their contributions comparable. Models were run in R (R Core Team, 2021) using the lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017) packages, and were optimized using the bobyqa optimizer. Effect sizes equivalent to Cohen's d for mixed-effects models were calculated using the EMAtools R package (Kleiman, 2021).

Results

Mouth-Looking in Song and Speech Across Development

Figure 3 shows infants' mouth-looking averages for ID song and speech during the first year of life, with the predictions from the first model (i.e., mouth-looking as a function of age, and clip type) overlaid. Infants looked progressively more at the mouth as they aged in both song and speech conditions ($B = 0.035$, $p < .001$, $d = 2.00$). However, mouth-looking was higher in song than speech stimuli overall ($B = 0.100$, $p < .001$, $d = 0.42$), and the increase in mouth-looking across development occurred faster for song than speech ($B = 0.010$, $p < .001$, $d = 0.12$). In this and the second model, there was a marginal main effect of sex on mouth-looking , with males fixating on the mouth less than females overall ($B = -0.035$, $p = .08$, $d = -0.21$). For the interested reader, additional analyses for overall attention (total fixations to all ROIs) to song and speech, which replicate findings of preferential attention to song in the first year of life (Nakata & Trehub, 2004; Trehub et al., 2016), can be found in the Supplementary Materials.

To assess how early mouth-looking increases in song relative to speech contexts, we conducted an additional analysis fitting the model only to data points in the first half of the first year of life (between 2.5 to 6.4 months of age). Mouth-looking trajectories for song and speech diverged early in development with an already significant increase in mouth-looking for song but not speech by 6.4 months of age ($B = 0.038$, $p < .001$, $d = 0.19$).

Note that higher mouth-looking in song versus speech does not necessarily mean preferential mouth-looking over eye-looking at all developmental time points in song. Younger infants in our sample still looked preferentially to the eyes in both speech and song, but this preference for eyes was reduced in song starting early in development, with an earlier and faster shifting of attention towards the mouth in song than speech.

Clip Characteristics as Drivers of Mouth-Looking Across Development

In the second model, we exchanged clip type for the clip features and their interactions with age. Our song and speech stimuli differed along multiple featural dimensions with song having slower tempo, lower rhythmic variability, increased salience of AVS in the mouth ROI, and increased positive affect (Table 1 and Figure 2). This feature-level model significantly improved model fit ($AIC = 576.22$) when compared with the initial model that considered only clip type and age ($AIC = 618.10$, $\chi^2(6) = 53.878$, $p < .001$). Tempo, rhythmicity, positive affect, AVS saliency in the mouth, and their interactions with age all improved model fit. Figure 4 shows the model predictions for each of the features across age points.

As in the previous model, older infants looked at the mouth more overall ($B = 0.031$, $p < .001$, $d = 1.86$). Infants also increased their overall mouth-looking for clips with slower tempo ($B = -0.042$, $p < .001$, $d = -0.30$), lower rhythmic variability ($B = -0.010$, $p < .01$, $d = -0.07$), higher positive affect ($B = 0.009$, $p < .05$, $d = 0.06$), and higher salience of AVS in the mouth as compared to the eyes ($B = 0.010$, $p < .01$, $d = 0.07$). The strength of these predictors, however, varied with age: slower tempo was a progressively stronger predictor of mouth-looking as infants aged ($B = -0.005$, $p < .001$, $d = -0.10$), as was lower rhythmic

variability ($B = -0.002$, $p < .05$, $d = -0.05$), and higher mouth AVS saliency ($B = 0.003$, $p < .01$, $d = 0.07$). In contrast, the effects of positive affect on mouth-looking decreased as infants got older ($B = -0.003$, $p < .01$, $d = -0.07$). However, we note these effects of the interactions between age and clip characteristics were generally quite small.

Discussion

ID speech and song are ubiquitous communicative signals in infants' daily environments (Mendoza & Fausey, 2022; Steinberg et al., 2021; Trehub et al., 1997; Yan et al., 2021). While both ID speech and song capture infants' attention, ID song is particularly effective at maintaining infants' overall attention particularly in combination with visual information from the singer's face (Costa-Giomi, 2014, Macari et al., 2021; Trehub et al., 2016), an effect replicated in the current study. Moreover, here we demonstrate that infants' attention allocation within an engaging face during infant-directed communication differed for ID song and ID speech during the first year of life. Engaging infants with ID song resulted in more infant mouth-looking than engaging them with ID speech. In line with previous studies (Hillaret de Boisferon et al., 2017; Lewkowicz & Hansen-Tift, 2012; Tenenbaum et al., 2013; Wagner et al., 2013), mouth-looking increased with age in both contexts, however, the shift from eyes to mouth started earlier and increased faster for song as compared to speech.

The increased mouth-looking during ID song versus ID speech is at least in part driven by the features that naturally vary across these communicative categories such as tempo, rhythmicity, audiovisual synchrony, and positive affect. Infants demonstrate early sensitivity to these types of amodal cues that are present in both the audio and visual aspects of multimodal signals (i.e., intersensory redundancy; Bahrick & Lickliter., 2000; Bahrick et al., 2004; Flom & Bahrick, 2007; Lewkowicz, 2003; Lewkowicz & Marcovitch, 2006). Some of these features have previously been demonstrated to impact mouth-looking using more constrained experimental stimuli, and their influence is now demonstrated here taking advantage of their natural variability in ecologically-valid stimuli. For example, in line with reports of increased fixation duration to the mouth during experimentally slowed-down speech in older children (Gepner et al., 2021), infants in the current study looked more at the mouth during clips with slower vocalization rates overall. Infants also looked more toward the mouth for clips with lower rhythmic variability, which corresponds with increased rhythmic predictability, and is a hallmark of sung interactions (Hilton et al., 2022; Savage et al., 2021; Trainor, 1996). We also observed increased mouth-looking with greater relative mouth AVS, consistent with the sensitivity to AVS noted in prior studies of experimentally desynchronized speech stimuli (Hillareit de Boisferon et al., 2017). Tempo, rhythmicity, and AVS all had stronger effects on mouth-looking *across* speech and song contexts in older infants (although note the effects of the interactions with age are quite subtle), suggesting sensitivity to these features for promoting mouth-looking may be most apparent during certain developmental periods when they could serve as mechanisms relevant for language learning. For example, the audiovisual synchrony that is exclusively available in the mouth movements (i.e., fine-grained articulatory information) would be more relevant later in the first year of life for infants' language skills (cf. Hillareit de Boisferon et al., 2017; Lewkowicz et al., 2015; Tenembaum et al., 2015). This possibility will have to be further

examined in the future looking directly at language outcomes in relation to mouth-looking (or mouth-looking trajectories) in song and speech at different developmental time points.

Positive affect also significantly predicted mouth-looking with infants attending to the mouth more during clips in which the actors smiled more. Vocal and visual positive affect attract infants' attention during both song and speech (Corbeil et al., 2013; Kim & Johnson 2013; Trehub et al., 2016) though the engaging visual component – such as increased smiling during live sung versus spoken interactions – appears to be particularly important (Costa-Giomi, 2014; Trehub et al., 2016). The current data suggests that when presented with audiovisual recordings of ID speech and song, the presence of smiling specifically increases infant attention to the interlocutor's mouth region. This is consistent with previous results showing that 8- and 12-month-old infants attend more to the mouth of an adult that is laughing, as opposed to crying or displaying a neutral expression (Ayneto & Sebastian-Galles, 2016). However, in contrast to the other features investigated here, positive affect was a stronger predictor of mouth-looking earlier in infancy (though as with the other features, we note the interaction with age was quite small).

Examining the different features separately identified specific drivers of visual attention to the mouth across song and speech. These features, however, occur within these two contexts in specific non-trivial combinations that make speech and song complex and meaningful integrated socio-communicative signals that are perceived as an integrated whole. The specific combination of features that generally occur in song – including slower tempo, higher rhythmicity, increased AVS, and increased smiling (Hilton et al., 2022; Trainor, 1996; Trehub et al., 2016) – make this communicative context particularly good at promoting mouth-looking across infancy. Our findings that tempo, rhythmicity, and AVS are most predictive of mouth-looking in older infants, whereas positive affect is most predictive in younger infants, suggests that the adaptive value of these features –and their natural potentiation during song– changes across development. Perhaps song shifts from a context primarily important for affect regulation and social bonding (Cirelli et al., 2020; Corbeil et al., 2016; Trainor, 1996) to a context that also carries useful information for language development, as infants reach a developmental stage of increased receptivity for language learning. In the second half of the first year of life, increased mouth-looking during ID speech is associated with children's concurrent and future expressive and receptive language skills (Imafuku & Myowa, 2016; Tenenbaum et al., 2015; Tsang et al., 2018; Young et al., 2009). Aspects of ID song such as rhythmic predictability and slow tempo, which highlight the suprasegmental features of speech, are theorized to serve as mechanisms underlying the potential role of song in language development (Falk et al., 2021; François et al., 2017; Jusczyk et al., 1999; Schön & François, 2011; Schön et al., 2008; Thiessen & Saffran, 2009). The current results suggest these attributes, embedded in a song context, may also support language learning via promoting attention to a singer's mouth, but such possibility needs to be investigated in future research.

Existing evidence does suggest that visual facial features support language skills across speech and song contexts. For example, neural tracking of auditory-only nursery rhymes in 10- and 14-month-olds predicts vocabulary size at 24 months (Menn et al., 2022). For speech stimuli, neural tracking is greater in both infants and adults when audiovisual

information from the speaker's face is available (Tan et al., 2022). Neural tracking is also greater in sung than spoken sentences in adults listening to audio only stimuli (der Nederlanden et al., 2020; der Nederlanden et al., 2022). Furthermore, in adults, seeing a singer's face increases lyric comprehension (Jesse & Massaro, 2010). That being said, however, given ID speech's strong and direct connection to language development (Golinkoff et al., 2015; Thiessen et al., 2005), and ID song's strong connection to attentional capture and affect regulation (Cirelli et al., 2020; Corbeil et al., 2016; Hilton et al., 2022; Trainor, 1996), it is possible that, in alignment with these proposed functions, the characteristics we explored in this study may show differential effects on mouth-looking, as well as its possible support of language skills, when studied separately *within* each of these contexts. As well, the utility of such features for speech and language skills may be most meaningful at different developmental time points for speech and song contexts. The current dataset does not support running separate analyses of clip characteristics within contexts, due to concerns regarding distribution of clip characteristics within categories for such analyses to be robust. Future studies should further explore this matter using a larger number of speech and song clips across age points, with full coverage of the feature space while still being naturalistic.

Overall increased infant attention to ID song over speech has been considered in line with the theory of ID song as a credible signal of parental attention (Mehr & Krasnow, 2017; Mehr et al., 2020). Increased attention to the mouth in song may perhaps also be in line with an extension of such theory into the visual components of the vocal signal. For example, the visual correlates of "infant-directedness" of an acoustic signal – such as its rhythmicity, repetitiveness, and elevated pitch (Hilton et al., 2022) – are most apparent in the mouth region of the caregiver (Chandrasekaran et al., 2009; Livingstone et al., 2015; Trainor, 1996; Trehub et al., 2016; Yehia et al., 2018). Thus, in a slower, more rhythmic, and more affectively positive signal like song, these amodal and visual features of infant-directedness may more readily lead the onlooking infant to attend to the mouth region via their increased multimodal redundancy, which is salient to infants (cf. Bahrick & Lickliter., 2000; Bahrick et al., 2004). Of course, in the youngest infants in the sample (< ~6 months), infants preferentially looked at the eyes in both speech and song contexts, consistent with the importance of eye-looking for social and emotional regulation (Farroni et al., 2002; Jones & Klin, 2013; Lense et al., 2022); however, even at these early age points, the preference for eyes was attenuated during song. Earlier and more rapid increases in mouth-looking in song may also reflect more efficient processing of the sung signal, perhaps in part due to these multimodal, infant-directed characteristics (e.g., slower tempo, increased AVS; Singh et al., 2009; Song et at., 2010). If this is the case, we may expect to see an earlier decline in mouth-looking in song than speech as a result of lower processing demands in song contexts. Future studies could examine mouth-looking across a longer developmental period as increases in mouth-looking for ID speech have been noted during the second year of life (and beyond), as well (de Boisferon et al, 2018; Morin-Lessard et al., 2019; Pons et al., 2015).

Infants' increased mouth-looking during song, and in relation to specific clip features that are more prominent in song, is of methodological relevance, as well. The current results highlight the importance of stimuli selection for studies exploring infants' fixation patterns

during infant-directed communication. While prior studies highlight the increase in mouth-looking to infant-directed speech over the first years of life, the precise developmental timing of preferences for mouth has varied some across studies (Frank et al., 2012; Hillaret de Boisferon, 2017; Morin-Lessard et al., 2019; Sekiyama et al., 2021; Tenenbaum et al., 2013). It is possible that some of the differences observed in the literature with respect to eyes/mouth-looking tradeoffs and their timing might be in part related to differences in stimuli features like the ones explored here (e.g., degree of rhythmicity or tempo of stimuli). Differences in stimuli selection and characteristics may also underlie some of the differences observed in mouth-looking trajectories during ID speech in the current study versus some prior reports in the literature, which suggest infant preferential attention to a speaker's mouth versus eyes by ~8 months of age (Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015).

In the present study, infants being engaged with ID speech increasingly looked at the mouth more as they aged, but they only reached preferential mouth-looking relative to eye-looking by 12 months of age. In addition to possible featural differences of stimuli, our study and stimuli differed from prior studies in several ways. Previous studies generally show only the face and neck of the actors, who are set against neutral/plain backgrounds. These studies typically employ only 1–2 actors who recite only a limited script (e.g., 1–2 monologues or stories). In contrast, our clips showed not only the face and neck but also some of the upper torso of the actors, and actors were set against a visually interesting background that resembled a nursery. The availability of additional intersensory redundancy in the body (as a result of body movements), as well as additional elements to attend to in the background, might have modulated infants' viewing patterns. Our stimuli also included more variability in numbers of actors, songs, and speech contexts than previous studies, and such increased variability may have shifted fixation patterns, as well. In this respect, the current stimuli reflect the rich variability infants will experience in everyday interactions, and make our results more generalizable to a wider set of contexts. Relatedly, while prior studies have used cross-sectional samples for different age points, our sample was longitudinal. While we believe this is a strength of the current study, enabling us to map developmental trajectories of mouth-looking in the speech and song contexts, this may have also impacted infants' attention allocation over time (e.g., even due to comfort in the lab setting (e.g., Santolin et al., 2021)).

Our results open several new avenues for future inquiry. First, studies to date have provided evidence of the relationship between infants' attention to the mouth during *ID speech* and their expressive and receptive language skills (Imafuku & Myowa, 2016; Tenenbaum et al., 2015; Tsang et al., 2018; Young et al., 2009). Future research should test whether this relationship also holds true for mouth-looking during *ID song*, and whether mouth-looking associated with specific features of song and speech, and/or at different points in the developmental trajectory is uniquely predictive of later language development. Second, future studies should also explore the influence of individual differences on song's potentiation of attention to the mouth. Infants are generally exposed repeatedly to a limited set of songs (Mendoza & Fausey, 2021), prefer familiar songs (Kragness et al., 2022), and increase their overall attention to others who sing familiar songs (Cirelli & Trehub, 2020). However, if or how specific song familiarity, or degree of exposure to song more

generally, moderates infants' attention allocation to a singer's face is unknown. Third, by increasing infants' attention to a singer's mouth, song might offer a tool to support language development in clinical populations (although see discussion of different functions of speech and song above) such as those with or at elevated likelihood of communication challenges (e.g., due to language impairment, autism, or hearing impairment). For example, attention to the mouth during audiovisual *ID speech* is associated with language acquisition in typically developing infants but not in children with elevated likelihood of developing autism (Chawarska et al., 2022). Future studies could additionally examine mouth-looking during *ID song* and language skills in this population, given that song is more rhythmic and predictable (Hilton et al., 2022), promotes attention to the face in autistic children (Macari et al., 2020; Thompson & Abel, 2016), and provides opportunities for social engagement (Lense & Camarata, 2020).

Our stimuli were designed with ecological validity in mind, and so they reflected the natural variability in contexts and communicative styles that children might be exposed to at different ages. This emphasis on naturalness allowed us to study infants' responses to the features of ID song and ID speech as they usually occur in these two contexts, which in turn supported more ecologically valid inferences on the possible mechanisms of ID communication that may support language learning. However, this methodological choice also meant reduced experimental control of the individual and joint distributions of clip features children were exposed to across time points. Future studies could experimentally manipulate the occurrence and co-occurrence of specific features within each context to further our understanding of their individual and combined contributions as drivers of attention to the mouth. Slowing down speech, for example, enhances young children's attention to the mouth (Gepner et al., 2021); perhaps speeding up song, or breaking its rhythmic pattern, might reduce attention to the mouth.

A potential limitation of the naturalness of our stimuli was that the specific clips shown to infants changed with development. While some clips were shown across multiple ages, other clips were varied across the playlists as children aged and were able to attend for longer periods of time. Additional analyses presented in the Supplemental Methods indicate that the clip characteristics remain generally stable over time. However, future studies could more purposefully repeat and vary clips over time in relationship to the clip characteristics to more fully investigate the featural analyses, particularly for considering changes in sensitivity to characteristics with infant age. Additionally, our stimuli were recorded portrayals rather than taken from live interactions with an infant present. This facilitated the recording of cleaner audio files, as well as performances that were not driven by any particular feedback from an infant (cf. Smith & Trainor, 2008). However, the presence versus absence of an infant changes the acoustic and visual characteristics of ID communication (Trehub et al, 1993; Trehub et al., 1997; Trehub et al., 2016). The distributions of the audiovisual features across our clips suggests that actors were generally able to replicate traditional differences between song and speech reported in the literature (Hilton et al., 2022; Trainor et al., 2015). Nevertheless, measuring infants' visual attention to song and speech recorded in the presence of an actual infant, or during live interactions with a caregiver or other adult, is another exciting avenue of future exploration.

Conclusion

ID song is highly effective at capturing and maintaining infants' attention with its visual cues playing an important role in engaging infants (Costa-Giomi, 2014; Macari et al., 2021; Trehub et al., 2016). The current study demonstrates that beyond its overall attentional capture effects, ID song promotes mouth-looking in infants to a higher degree than ID speech. This effect is especially prominent during the latter half of the first year, a developmental period associated with increased language learning sensitivity. Song as a communicative context naturally combines many features that increase attention to the mouth during the first year of life: slower tempo, increased rhythmicity, increased audiovisual synchrony, and increased positive affect. Future studies can investigate whether, and at what developmental time points, ID song's modulation of infants' visual attention to a singer's mouth might provide a mechanism for supporting language learning in typically developing infants and infants with communication challenges.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work is supported by NIMH and NCCIH [R61MH123029], NIMH P50 [MH100029], NIDCD [R21DC016710], NICHD [R00HD097290], and ACM Lifting Lives Foundation. The authors would like to thank the children and families who participated in this study, as well as the eye-tracking labs, staff members, and clinicians at Marcus Autism Center who made this work possible.

Data availability statement

The data that support the findings of this study is openly available through OSF in this link: https://osf.io/qu5nz/

References

Ayneto A, & Sebastian-Galles N (2017). The influence of bilingualism on the preference for the mouth region of dynamic faces. Developmental Science, 20(1), e12446.

Bahrick LE, & Lickliter R (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. Developmental psychology, 36(2), 190. [PubMed: 10749076]

Bahrick LE, Lickliter R, & Flom R (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. Current Directions in Psychological Science, 13(3), 99–102.

Bahrick LE, McNew ME, Pruden SM, & Castellanos I (2019). Intersensory redundancy promotes infant detection of prosody in infant-directed speech. Journal of experimental child psychology, 183, 295–309. [PubMed: 30954804]

Bahrick LE, Walker AS, & Neisser U (1981). Selective looking by infants. Cognitive Psychology, 13(3), 377–390. [PubMed: 7237992]

Baltrušaitis T, Mahmoud M, & Robinson P (2015, May). Cross-dataset learning and person-specific normalisation for automatic action unit detection. In 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG) (Vol. 6, pp. 1–6). IEEE.

Baltrusaitis T, Zadeh A, Lim YC, & Morency LP (2018, May). Openface 2.0: Facial behavior analysis toolkit. In 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018) (pp. 59–66). IEEE.

Barenholtz E, Mavica L, & Lewkowicz DJ (2016). Language familiarity modulates relative attention to the eyes and mouth of a talker. Cognition, 147, 100–105. [PubMed: 26649759]

Bates D, Maechler M, Bolker B, & Walker S (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1–48. doi:10.18637/jss.v067.i01

Bergeson TR, & Trehub SE (2002). Absolute pitch and tempo in mothers' songs to infants. Psychological Science, 13(1), 72–75. [PubMed: 11892783]

Boersma P & Weenink D (2022). Praat: doing phonetics by computer [Computer program].

Brooks R, & Meltzoff AN (2002). The importance of eyes: how infants interpret adult looking behavior. Developmental psychology, 38(6), 958. [PubMed: 12428707]

Brooks R, & Meltzoff AN (2005). The development of gaze following and its relation to language. Developmental science, 8(6), 535–543. [PubMed: 16246245]

Bryant GA, & Barrett HC (2007). Recognizing intentions in infant-directed speech: Evidence for universals. Psychological Science, 18(8), 746–751. [PubMed: 17680948]

Buchan JN, Paré M, & Munhall KG (2007). Spatial statistics of gaze fixations during dynamic face processing. Social Neuroscience, 2(1), 1–13. [PubMed: 18633803]

Chandrasekaran C, Trubanova A, Stillittano S, Caplier A, and Ghazanfar AA (2009). The natural statistics of audiovisual speech. PLoS Comput. Biol. 5:e1000436. doi: 10.1371/journal.pcbi.1000436 [PubMed: 19609344]

Chawarska K, Lewkowicz D, Feiner H, Macari S, & Vernetti A (2022). Attention to audiovisual speech does not facilitate language acquisition in infants with familial history of autism. Journal of Child Psychology and Psychiatry.

Chong SCF, Werker JF, Russell JA, & Carroll JM (2003). Three facial expressions mothers direct to their infants. Infant and Child Development: An International Journal of Research and Practice, 12(3), 211–232.

Cirelli LK, Jurewicz ZB, & Trehub SE (2020). Effects of maternal singing style on mother–infant arousal and behavior. Journal of cognitive neuroscience, 32(7), 1213–1220. [PubMed: 30912725]

Cirelli LK, & Trehub SE (2020). Familiar songs reduce infant distress. Developmental psychology, 56(5), 861. [PubMed: 32162936]

Corbeil M, Trehub SE, & Peretz I (2013). Speech vs. singing: Infants choose happier sounds. Frontiers in psychology, 4, 372. [PubMed: 23805119]

Corbeil M, Trehub SE, & Peretz I (2016). Singing delays the onset of infant distress. Infancy, 21(3), 373–391.

Costa-Giomi E (2014). Mode of presentation affects infants' preferential attention to singing and speech. Music Perception: An Interdisciplinary Journal, 32(2), 160–169.

Costa-Giomi E, & Ilari B (2014). Infants' preferential attention to sung and spoken stimuli. Journal of Research in Music Education, 62(2), 188–194.

de la Cruz-Pavía I, Gervain J, Vatikiotis-Bateson E, & Werker JF (2020). Coverbal speech gestures signal phrase boundaries: A production study of Japanese and English infant-and adult-directed speech. Language Acquisition, 27(2), 160–186.

der Nederlanden CMVB, Joanisse MF, & Grahn JA (2020). Music as a scaffold for listening to speech: Better neural phase-locking to song than speech. NeuroImage, 214, 116767. [PubMed: 32217165]

der Nederlanden CMVB, Joanisse MF, Grahn JA, Snijders TM, & Schoffelen JM (2022). Familiarity modulates neural tracking of sung and spoken utterances. NeuroImage, 252, 119049. [PubMed: 35248707]

Falk S, & Audibert N (2021). Acoustic signatures of communicative dimensions in codified mother-infant interactions. The Journal of the Acoustical Society of America, 150(6), 4429–4437. [PubMed: 34972287]

Falk S, Fasolo M, Genovese G, Romero-Lauro L, & Franco F (2021). Sing for me, Mama! Infants' discrimination of novel vowels in song. Infancy, 26(2), 248–270. [PubMed: 33523572]

Farroni T, Csibra G, Simion F, & Johnson MH (2002). Eye contact detection in humans from birth. Proceedings of the National academy of sciences, 99(14), 9602–9605.

Fernald A (1985). Four-month-old infants prefer to listen to motherese. Infant behavior and development, 8(2), 181–195.

Fernald A (1989). Intonation and communicative content in mothers' speech to infants: Is the melody the message? Child Development, 60, 1497–1510. doi:10.2307/1130938 [PubMed: 2612255]

Fernald A, Taeschner T, Dunn J, Papoušek M, de Boysson-Bardies B, Fukui I (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. Journal of Child Language, 16, 477–501. [PubMed: 2808569]

Flom R, & Bahrick LE (2007). The development of infant discrimination of affect in multimodal and unimodal stimulation: The role of intersensory redundancy. Developmental psychology, 43(1), 238. [PubMed: 17201522]

Franco F, Suttora C, Spinelli M, Kozar I, & Fasolo M (2021). Singing to infants matters: early singing interactions affect musical preferences and facilitate vocabulary building. Journal of Child Language, 1–26.

François C, Teixidó M, Takerkart S, Agut T, Bosch L, & Rodriguez-Fornells A (2017). Enhanced neonatal brain responses to sung streams predict vocabulary outcomes by age 18 months. Scientific reports, 7(1), 1–13. [PubMed: 28127051]

Frank MC, Vul E, & Saxe R (2012). Measuring the development of social attention using free-viewing. Infancy, 17(4), 355–375. [PubMed: 32693486]

Gepner B, Godde A, Charrier A, Carvalho N, & Tardif C (2021). Reducing facial dynamics' speed during speech enhances attention to mouth in children with autism spectrum disorder: An eye-tracking study. Development and Psychopathology, 33(3), 1006–1015. [PubMed: 32378498]

Gerry D, Unrau A, & Trainor LJ (2012). Active music classes in infancy enhance musical, communicative and social development. Developmental Science, 15(3), 398–407. [PubMed: 22490179]

Golinkoff RM, Can DD, Soderstrom M, & Hirsh-Pasek K (2015). (Baby) talk to me: the social context of infant-directed speech and its effects on early language acquisition. Current Directions in Psychological Science, 24(5), 339–344.

Grieser DL, & Kuhl PK (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. Developmental psychology, 24(1), 14.

Hannon EE, Lévêque Y, Nave KM, & Trehub SE (2016). Exaggeration of language-specific rhythms in English and French children's songs. Frontiers in Psychology, 7, 939. [PubMed: 27445907]

Hillairet de Boisferon A, Tift AH, Minar NJ, & Lewkowicz DJ (2017). Selective attention to a talker's mouth in infancy: role of audiovisual temporal synchrony and linguistic experience. Developmental Science, 20(3), e12381.

Hillareit de Boisferon AH, Tift AH, Minar NJ, & Lewkowicz DJ (2018). The redeployment of attention to the mouth of a talking face during the second year of life. Journal of Experimental Child Psychology, 172, 189–200. [PubMed: 29627481]

Hilton CB, Moser CJ, Bertolo M, Lee-Rubin H, Amir D, Bainbridge CM, ... & Mehr SA. (2022). Acoustic regularities in infant-directed speech and song across cultures. Nature Human Behaviour, 1–12.

Imafuku M, & Myowa M (2016). Developmental change in sensitivity to audiovisual speech congruency and its relation to language in infants. Psychologia, 59(4), 163–172.

Jesse A, & Massaro DW (2010). Seeing a singer helps comprehension of the song's lyrics. Psychonomic bulletin & review, 17(3), 323–328. [PubMed: 20551353]

Jones W, & Klin A (2013). Attention to eyes is present but in decline in 2–6-month-old infants later diagnosed with autism. Nature, 504(7480), 427–431. [PubMed: 24196715]

Jusczyk PW, Houston DM, & Newsome M (1999). The beginnings of word segmentation in English-learning infants. Cognitive psychology, 39(3–4), 159–207. [PubMed: 10631011]

Karlsson S (2022). Optical Flow with Matlabs Computer vision toolbox (https://www.mathworks.com/matlabcentral/fileexchange/44611-optical-flow-with-matlabs-computer-vision-toolbox), MATLAB Central File Exchange. Retrieved April 4, 2022.

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

Kleiman E (2021). EMAtools: Data management tools for real-time monitoring/ecological momentary assessment data. R package version 0.1.4. https://CRAN.R-project.org/package=EMAtools

Kragness HE, Johnson EK, & Cirelli LK (2022). The song, not the singer: Infants prefer to listen to familiar songs, regardless of singer identity. Developmental Science, 25(1), e13149. [PubMed: 34241934]

Lartillot O, Toiviainen P, & Eerola T (2008). A matlab toolbox for music information retrieval. In Preisach C, Burkhardt H, Schmidt-Thieme L, Decker R (Eds.), Data analysis, machine learning and applications (pp. 261–268). Springer:Berlin.

Lense MD, & Camarata S (2020). PRESS-play: Musical engagement as a motivating platform for social interaction and social play in young children with ASD. Music & science, 3, 2059204320933080.

Lense MD, Shultz S, Astésano C, & Jones W (2022). Music of infant-directed singing entrain infants' social visual behavior. Proceedings of the National Academy of Sciences

Lewkowicz DJ (2003). Learning and discrimination of audiovisual events in human infants: the hierarchical relation between intersensory temporal synchrony and rhythmic pattern cues. Developmental psychology, 39(5), 795. [PubMed: 12952394]

Lewkowicz DJ, & Hansen-Tift AM (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. Proceedings of the National Academy of Sciences, 109(5), 1431–1436.

Lewkowicz DJ, & Marcovitch S (2006). Perception of audiovisual rhythm and its invariance in 4-to 10-month-old infants. Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology, 48(4), 288–300.

Lewkowicz DJ, Minar NJ, Tift AH, & Brandon M (2015). Perception of the multisensory coherence of fluent audiovisual speech in infancy: Its emergence and the role of experience. Journal of experimental child psychology, 130, 147–162. [PubMed: 25462038]

Livingstone SR, Thompson WF, Wanderley MM, & Palmer C (2015). Common cues to emotion in the dynamic facial expressions of speech and song. Quarterly Journal of Experimental Psychology, 68(5), 952–970.

Macari S, Milgramm A, Reed J, Shic F, Powell KK, Macris D, & Chawarska K (2021). Context-specific dyadic attention vulnerabilities during the first year in infants later developing autism spectrum disorder. Journal of the American Academy of Child & Adolescent Psychiatry, 60(1), 166–175. [PubMed: 32061926]

Mehr SA, & Krasnow MM (2017). Parent-offspring conflict and the evolution of infant-directed song. Evolution and Human Behavior, 38(5), 674–684.

Mehr SA, Krasnow MM, Bryant GA, & Hagen EH (2021). Origins of music in credible signaling. Behavioral and Brain Sciences, 44.

Mendoza JK, & Fausey CM (2021). Everyday music in infancy. Developmental Science, 24(6), e13122. [PubMed: 34170059]

Menn KH, Ward EK, Braukmann R, Van den Boomen C, Buitelaar J, Hunnius S, & Snijders TM (2022). Neural tracking in infancy predicts language development in children with and without family history of autism. Neurobiology of Language, 3(3), 495–514. [PubMed: 37216063]

Morin-Lessard E, Poulin-Dubois D, Segalowitz N, & Byers-Heinlein K (2019). Selective attention to the mouth of talking faces in monolinguals and bilinguals aged 5 months to 5 years. Developmental Psychology, 55(8), 1640. [PubMed: 31169400]

Nakata T, & Trehub SE (2004). Infants' responsiveness to maternal speech and singing. Infant Behavior and Development, 27(4), 455–464.

Papadimitriou A, Smyth C, Politimou N, Franco F, & Stewart L (2021). The impact of the home musical environment on infants' language development. Infant Behavior and Development, 65, 101651. [PubMed: 34784522]

Papousek M, Papousek H, & Symmes D (1991). The meanings of melodies in motherese in tone and stress languages. Infant Behavior & Development, 14, 415–440. doi:10.1016/0163-6383(91)90031-M

Patel AD, Iversen JR, & Rosenberg JC (2006). Comparing the rhythm and melody of speech and music: The case of British English and French. The Journal of the Acoustical Society of America, 119(5), 3034–3047. [PubMed: 16708959]

Pons F, Bosch L, & Lewkowicz DJ (2015). Bilingualism modulates infants' selective attention to the mouth of a talking face. Psychological science, 26(4), 490–498. [PubMed: 25767208]

Pons F, Bosch L, & Lewkowicz DJ (2019). Twelve-month-old infants' attention to the eyes of a talking face is associated with communication and social skills. Infant Behavior and Development, 54, 80–84. [PubMed: 30634137]

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Savage PE, Brown S, Sakai E, & Currie TE (2015). Statistical universals reveal the structures and functions of human music. Proceedings of the National Academy of Sciences, 112(29), 8987–8992.

Schmidt KL, & Cohn JF (2001, August). Dynamics of facial expression: Normative characteristics and individual differences. In IEEE International Conference on Multimedia and Expo, 2001. ICME 2001. (pp. 547–550). IEEE.

Schön D, Boyer M, Moreno S, Besson M, Peretz I, & Kolinsky R (2008). Songs as an aid for language acquisition. Cognition, 106(2), 975–983. [PubMed: 17475231]

Schön D, & François C (2011). Musical expertise and statistical learning of musical and linguistic structures. Frontiers in psychology, 2, 167. [PubMed: 21811482]

Sekiyama K, Hisanaga S, & Mugitani R (2021). Selective attention to the mouth of a talker in Japanese-learning infants and toddlers: Its relationship with vocabulary and compensation for noise. Cortex, 140, 145–156. [PubMed: 33989900]

Singh L, Nestor S, Parikh C, & Yull A (2009). Influences of infant-directed speech on early word recognition. Infancy, 14(6), 654–666. [PubMed: 32693515]

Smith NA, & Trainor LJ (2008). Infant-directed speech is modulated by infant feedback. Infancy, 13(4), 410–420.

Song JY, Demuth K, & Morgan J (2010). Effects of the acoustic properties of infant-directed speech on infant word recognition. The Journal of the Acoustical Society of America, 128(1), 389–400. [PubMed: 20649233]

Steinberg S, Shivers CM, Liu T, Cirelli LK, & Lense MD (2021). Survey of the home music environment of children with various developmental profiles. Journal of Applied Developmental Psychology, 75, 101296. [PubMed: 34737486]

Stern DN (1974). Mother and infant at play: the dyadic interaction involving facial, vocal, and gaze behaviors. In Lewis M & Rosenblum L (Eds.), The effect of the infant on its caregiver (pp. 187–232). New York: Wiley.

Stern DN, Spieker S, Barnett RK, & MacKain K (1983). The prosody of maternal speech: Infant age and context related changes. Journal of Child Language, 10(1), 1–15. [PubMed: 6841483]

Swerts M, & Krahmer E (2008). Facial expression and prosodic prominence: Effects of modality and facial area. Journal of Phonetics, 36(2), 219–238.

Symons LA, Hains SM, & Muir DW (1998). Look at me: Five-month-old infants' sensitivity to very small deviations in eye-gaze during social interactions. Infant Behavior and Development, 21(3), 531–536.

Tan SJ, Kalashnikova M, Di Liberto GM, Crosse MJ, & Burnham D (2022). Seeing a Talking Face Matters: The Relationship between Cortical Tracking of Continuous Auditory-Visual Speech and Gaze Behaviour in Infants, Children and Adults. NeuroImage, 119217. [PubMed: 35436614]

Tartter VC (1980). Happy talk: Perceptual and acoustic effects of smiling on speech. Perception & psychophysics, 27(1), 24–27. [PubMed: 7367197]

Tenenbaum EJ, Shah RJ, Sobel DM, Malle BF, & Morgan JL (2013). Increased focus on the mouth among infants in the first year of life: A Longitudinal Eye-Tracking Study. Infancy, 18(4), 534–553. [PubMed: 23869196]

Tenenbaum EJ, Sobel DM, Sheinkopf SJ, Malle BF, & Morgan JL (2015). Attention to the mouth and gaze following in infancy predict language development. Journal of Child Language, 42(6), 1173–1190. [PubMed: 25403090]

Thiessen ED, Hill EA, & Saffran JR (2005). Infant-directed speech facilitates word segmentation. Infancy, 7(1), 53–71. [PubMed: 33430544]

Thiessen ED, & Saffran JR (2009). How the melody facilitates the message and vice versa in infant learning and memory. Annals of the New York Academy of Sciences, 1169(1), 225–233. [PubMed: 19673786]

Thompson GA, & Abel LA (2018). Fostering Spontaneous Visual Attention in Children on the Autism Spectrum: A Proof-of-Concept Study Comparing Singing and Speech. Autism Research, 11(5), 732–737. [PubMed: 29356417]

Tomasello M, Hare B, Lehmann H, & Call J (2007). Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eye hypothesis. Journal of Human Evolution, 52(3), 314–320. [PubMed: 17140637]

Trainor LJ (1996). Infant preferences for infant-directed versus noninfant-directed playsongs and lullabies. Infant Behavior and Development, 19(1), 83–92.

Trehub SE, Ghazban N, Corbeil M (2015). Musical affect regulation in infancy. Annals of the New York Academy of Sciences, 1337, 186–192. [PubMed: 25773634]

Trehub SE, Plantinga J, & Russo FA (2016). Maternal vocal interactions with infants: Reciprocal visual influences. Social Development, 25(3), 665–683.

Trehub SE, Unyk AM, Kamenetsky SB, Hill DS, Trainor LJ, Henderson JL, & Saraza M (1997). Mothers' and fathers' singing to infants. Developmental Psychology, 33(3), 500. [PubMed: 9149928]

Trehub SE, Unyk AM, & Trainor LJ (1993). Maternal singing in cross-cultural perspective. Infant Behavior and Development, 16(3), 285–295.

Tsang T, Atagi N, & Johnson SP (2018). Selective attention to the mouth is associated with expressive language skills in monolingual and bilingual infants. Journal of experimental child psychology, 169, 93–109. [PubMed: 29406126]

Tsang CD, Falk S, & Hessel A (2017). Infants prefer infant-directed song over speech. Child development, 88(4), 1207–1215. [PubMed: 27796032]

Wagner JB, Luyster RJ, Yim JY, Tager-Flusberg H, & Nelson CA (2013). The role of early visual attention in social development. International journal of behavioral development, 37(2), 118–124. [PubMed: 26478642]

Weide RL, 1998. The Carnegie Mellon pronouncing dictionary. <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>.

Werker JF, & McLeod PJ (1989). Infant preference for both male and female infant-directed talk: a developmental study of attentional and affective responsiveness. Canadian Journal of Psychology/Revue canadienne de psychologie, 43(2), 230.

Werker JF, Pegg JE, & McLeod PJ (1994). A cross-language investigation of infant preference for infant-directed communication. Infant Behavior and Development, 17(3), 323–333.

Yan R, Jessani G, Spelke ES, de Villiers P, de Villiers J, & Mehr SA (2021). Across demographics and recent history, most parents sing to their infants and toddlers daily. Philosophical Transactions of the Royal Society B, 376(1840), 20210089.

Young GS, Merin N, Rogers SJ, & Ozonoff S (2009). Gaze behavior and affect at 6 months: predicting clinical outcomes and language development in typically developing infants and infants at risk for autism. Developmental science, 12(5), 798–814. [PubMed: 19702771]

Zadeh A, Chong Lim Y, Baltrusaitis T, & Morency LP (2017). Convolutional experts constrained local model for 3d facial landmark detection. In Proceedings of the IEEE International Conference on Computer Vision Workshops (pp. 2519–2528).

Research Highlights

- Infants' visual attention to adults' mouths during infant-directed speech has been found to support speech and language development.
- Infant-directed (ID) song promotes mouth-looking by infants to a greater extent than does ID speech across the first year of life.
- Features characteristic of ID song such as slower tempo, increased rhythmicity, increased audiovisual synchrony, and increased positive affect, all increase infants' attention to the mouth.
- The effects of song on infants' attention to the mouth are more prominent during the second half of the first year of life.

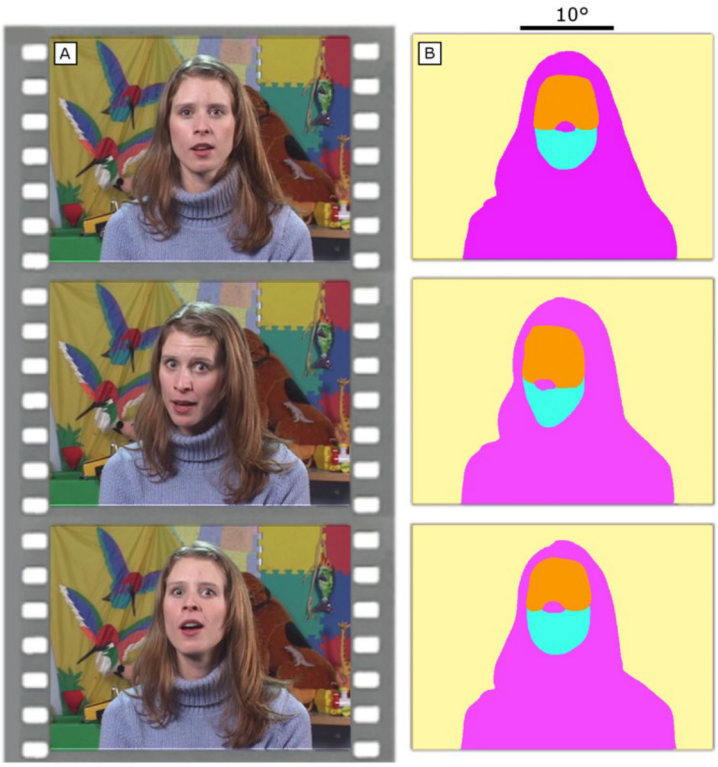

Figure 1.
Example of (A) still images and (B) corresponding regions of interest (ROIs: eyes: orange; mouth: cyan; body: fuchsia, and object: yellow) from a video clip in the current study. Our analyses focused on the proportion of face looking time spent fixating on the mouth ROI (PFLT-m: cyan ROI/(cyan ROI+ orange ROI)).

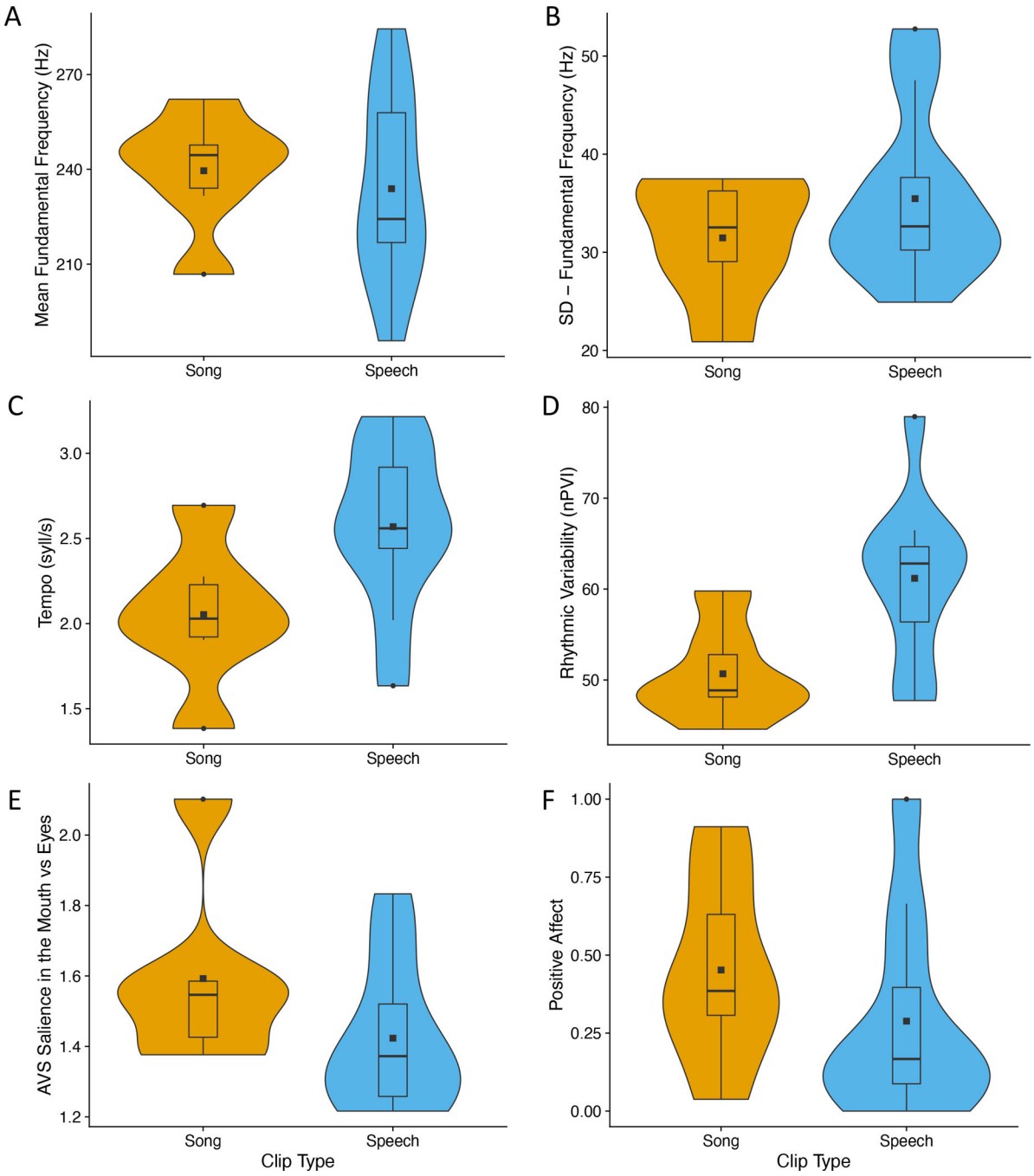

Figure 2.

Clip feature distributions for infant-directed song (orange - left) and infant-directed speech (blue - right): (A) Mean pitch; (B) Pitch variability (measured as standard deviation); (C) Tempo; (D) Rhythmic Variability; (E) Saliency of Audiovisual Synchrony (AVS) in the Mouth ROI; (F) Positive Affect. The grey squares show each feature's average per clip type.

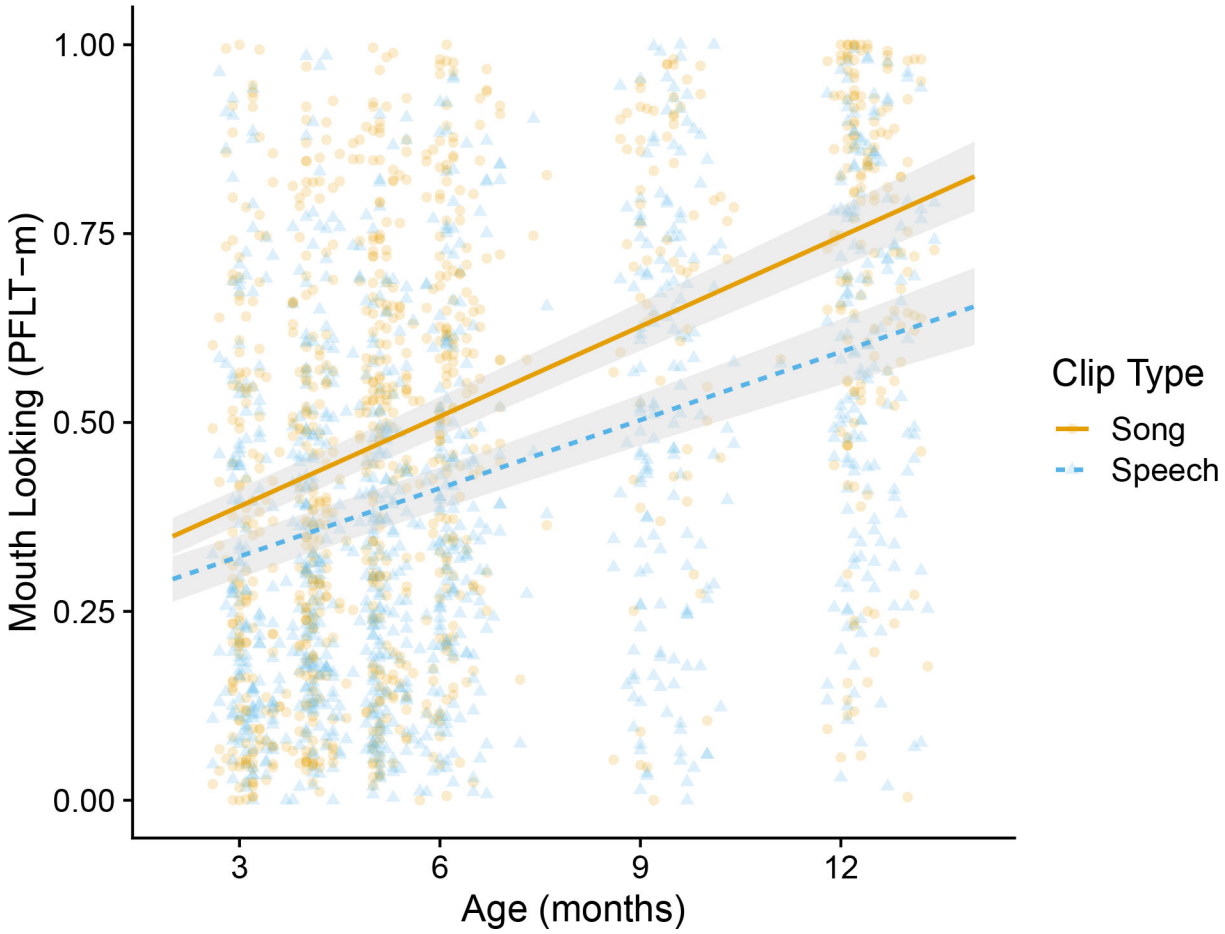

Figure 3.

Mouth-looking to infant-directed song and speech as a function of age. Mouth-looking is quantified as the percentage of face-looking time spent on the mouth region (PFLT-m). Note this measure is the complementary percentage of eye-looking and can be reversed and read with respect to attention to eyes (i.e., 60% mouth-looking can also be interpreted as 40% eye-looking). The individual points show average mouth-looking time per child at each age point for song (orange circles) and speech (blue triangles), and the lines show the model predictions for each clip type (song: orange solid; speech: blue dashed). The shaded regions represent 95% confidence intervals around model predictions.

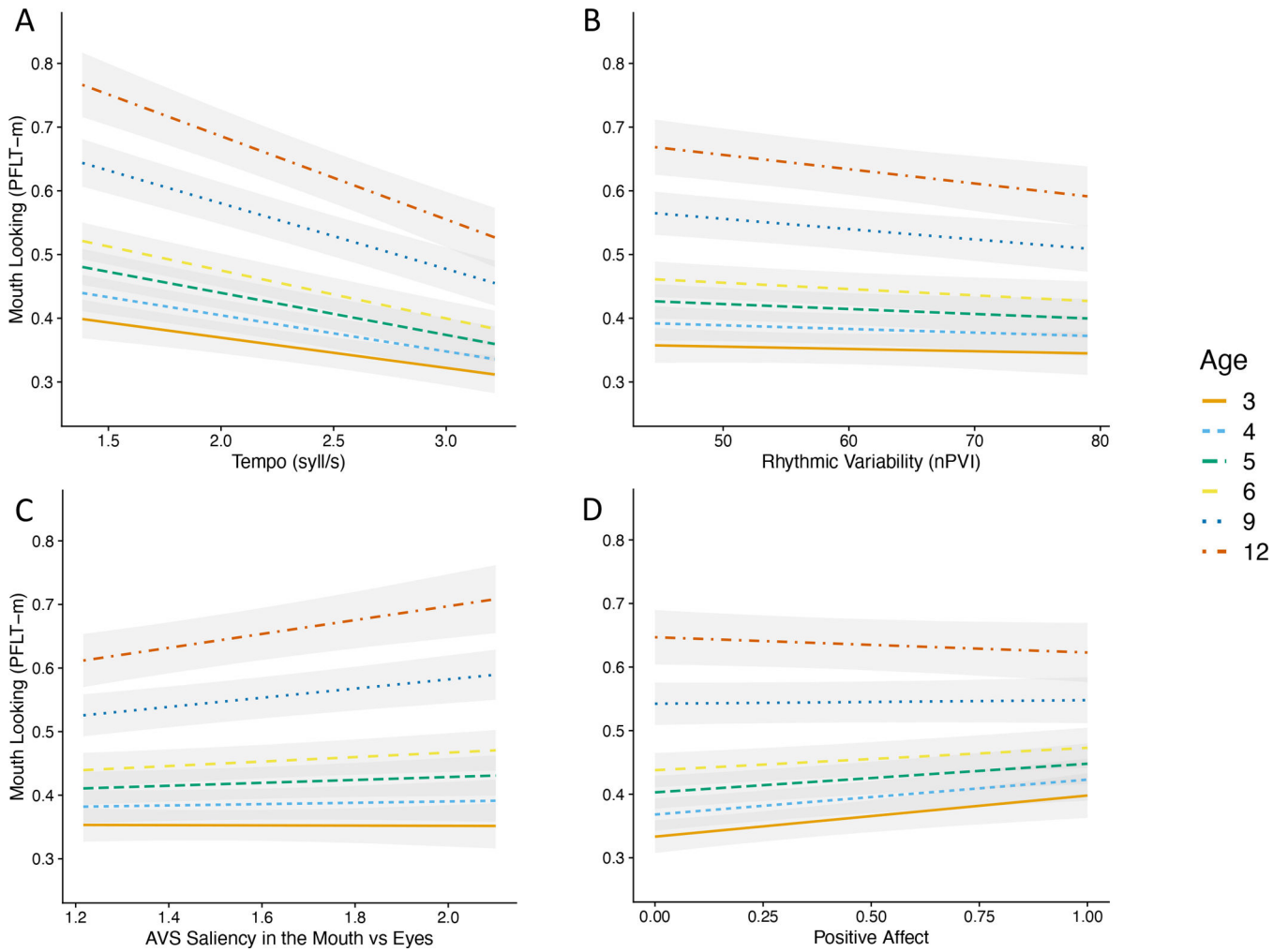

Figure 4.

Model predictions of mouth-looking (PFLT-m) as a function of clip features across age groups (indexed by color and line type): Mouth-looking as a function of (A) tempo; (B) rhythmic variability; (C) saliency of audiovisual synchrony (AVS) in the mouth ROI; (D) positive affect. Slower tempo, reduced rhythmic variability, and increased mouth AVS saliency all increased mouth-looking more in older infants while greater positive affect increased mouth-looking more for younger infants. The shaded regions represent 95% confidence intervals around the model predictions.

Table 1:

Summary of Clip Features for ID Song and ID Speech

Feature	Song M(SD)	Speech M(SD)	Mann-Whitney Test	Levene's Test
Pitch (Hz) – Mean	239.5 (18.9)	233.9 (30.2)	$U = 82.00$ $p = 0.792$	$F = 2.64$ $p = 0.13$
Pitch (Hz) – Standard Deviation	31.5 (6.4)	35.5 (8.7)	$U = 92.00$ $p = 0.492$	$F = 0.46$ $p = 0.51$
Tempo (Syllables/Second)	2.1 (0.4)	2.6 (0.5)	$U = 103.00$ $p = 0.056$	$F = 0.15$ $p = 0.70$
Rhythmic Variability (nPVI)	50.7 (5.4)	61.2 (9.1)	$U = 106.00$ $p = 0.022^*$	$F = 0.90$ $p = 0.36$
Saliency of Mouth AVS	1.6 (0.3)	1.4 (0.2)	$U = 42.00$ $p = 0.220$	$F = 0.00$ $p = 0.98$
Positive Affect (% Smiling)	45.2% (31.0%)	28.8% (32.3%)	$U = 75.00$ $p = 0.313$	$F = 0.04$ $p = 0.85$
Duration (Seconds)	23.0 (3.6)	18.5 (4.3)	$U = 92.00$ $p = 0.041^*$	$F = 3.06$ $p = 0.09$

Note:

*p <.05. Mann-Witney and Levene tests are provided as a reference of differences in averages and variance across clip types for the interested reader, but given the low number of clips, these statistical tests should be interpreted with caution as we may be underpowered to detect significant differences.