Sing for me, Mama!

Infants' discrimination of novel vowels in song

Falk, Simone[1,2]; Fasolo, Mirco[3]; Genovese, Giuliana[4]; Romero-Lauro, Leonor[4,5] & Franco, Fabia[6]

[1]Department of Linguistics and Translation, University of Montreal, Canada
[2]Laboratory for Brain, Music and Sound Research (BRAMS), University of Montreal, Canada
[3] Department of Neuroscience, Imaging, and Clinical Sciences, University "G. d'Annunzio" of Chieti-Pescara, Italy
[4]Department of Psychology, University of Milan-Bicocca, Milano, Italy
[5]Neuromi, Milan Center for Neuroscience, Milano, Italy
[6]Department of Psychology, Faculty of Science and Technology, Middlesex University London, UK

Conflict of interest statement: The authors do not have a conflict of interest to declare

Acknowledgments: Preliminary conceptual work to this study was funded by a grant from the Deutsche Forschungsgemeinschaft (DFG; program: initiation and enhancement of bilateral cooperation UK-Germany). The authors do not have a conflict of interest to declare.

Keywords: *infant-directed communication, singing, speech development, phonetic discrimination*

Abstract

When adults speak or sing with infants, they sound differently than in adult communication. Infant-directed (ID) communication helps caregivers to regulate infants' emotions, and helps infants to process speech information, at least from ID-*speech*. However, it is largely unclear whether infants might also process speech information presented in ID-*singing*. Therefore, we examined whether infants discriminate vowels in ID-singing, as well as potential differences with ID-speech. Using an Alternating-Trial-Preference-Procedure, infants aged 4-6 and 8-10 months were tested on their discrimination of an unfamiliar non-native vowel contrast presented in ID-like speech and singing. Relying on models of early speech sound perception, we expected that infants in their first half year of life would discriminate the vowels, in contrast to older infants whose non-native sound perception should deteriorate, at least in ID-like speech. Our results showed that infants of both age groups were able to discriminate the vowels in ID-like singing, while only the younger group discriminated the vowels in ID-like speech. These results show that infants process speech sound information in song from early on. They also hint at diverging perceptual or attentional mechanisms guiding infants' sound processing in ID-speech vs. -singing towards the end of the first year of life.

When adults address infant listeners, their speech is clearly different than when they speak with adults. Characteristic acoustic modifications distinguish infant-directed (ID) from adult-directed (AD) speech. Caregivers of infants speak with higher pitch, use a larger pitch range and display an ID-specific set of pitch contours (e.g., Fernald, 1989; Fernald et al., 1989; Fernald & Simon, 1984; Liu et al., 2007). They expand vowel sounds and use more distinct vowels compared to AD-speech (e.g., Burnham et al., 2002; Hartman et al., 2017; Kuhl et al., 1997; Weirich & Simpson, 2019). They also Utterances are shorter, pauses are longer and more frequent in ID-speech and pre-boundary vowels and syllables are lengthened (Albin & Echols, 1996; Bernstein Ratner, 1986; Fernald et al., 1989; Ludusan et al., 2016; Martin et al., 2016). Spectral modifications include a more emotional, happy-sounding tone of voice comparable to what adults use in romantic relationships (Bombar & Littig, 1996; Fernald, 1989; Piazza et al., 2017; Trehub et al., 2016). Overall, these acoustic modifications attract infants' attention to ID- vs. AD-speech (e.g., Cooper & Aslin, 1990; Soderstrom, 2019). They communicate caregivers' emotional availability (e.g., Bornstein et al., 2011), foster early caregiver-infant interaction (e.g., Gratier et al., 2015; Senju & Csibra, 2008), and can even impact infants' physical well-being (e.g., in preterm infants, Filippa et al., 2013). Higher amounts of ID-speech that infants receive in dyadic interactions with their caregivers were found to promote infants' general language development (e.g., lexical development, Ramírez-Esparza et al., 2014; Weisleder & Fernald, 2013).

ID-speech is not the only register used in early ID-communication. In fact, it is often produced on a continuum with ID-singing (Van Puyvelde & Franco, 2015) which forms an integral part of child rearing (e.g., Trehub, Becker, et al., 2015). Caregivers around the world frequently sing lullabies and playsongs to soothe and engage infants in early interactions (Trehub & Trainor, 1998). Thereby, they effectively modulate their arousal level and emotional state (Cirelli et al., 2020; Cirelli & Trehub, 2020; Corbeil et al., 2016). As the vast amount of ID-singing comes with lyrics set to a tune, infants are exposed to an additional source of language input (e.g., comprising phonetic/phonological, lexical, & morpho-syntactic information). Acoustically, ID-singing displays similar alterations compared to AD-singing as ID- vs. AD-speech. Caregivers prolong vowels, use higher pitch register in ID-singing, a loving tone of voice and an inventory of melodic pitch contours which is similar to ID-speech (e.g., Falk, 2011; Trainor et al., 1997; Trehub et al., 1993).

Besides the multiple benefits of ID-speech and -singing in early infancy in modulating infants' attention and arousal, fostering social and emotional development, evidence is scarce about the effects of both ID-registers on infants' early speech perception. During their first year of life, infants make immense progress in learning about the sounds of speech (Kuhl et al., 2005). Long before they start producing words, they have acquired fine-grained perceptual information about the characteristics of speech sounds. According to the 'perceptual reorganization' model (Werker & Tees, 1984), speech sound perception undergoes significant changes during the first year of life.

Infants in their first months of life have the ability to discriminate many speech sounds, across languages. For example, a classical study of Werker and colleagues (Werker et al., 1981) showed that 6-8 months old English-learning infants were able to discriminate two phonemic Hindi plosives (i.e., /t/-stops with retroflex vs. dental place of articulation), a contrast which is not present in English. Towards the end of the first year of life, 10-12 months old English-learning infants, like adults (Pruitt et al., 2006), no longer showed discrimination of the non-native contrast (Werker & Tees, 1984). For vowels, a similar reorganization is observed, which is supposed to start a little earlier (e.g., Kuhl et al., 1992; Polka & Werker, 1994). For example, English-learning infants discriminated the German front vs. back vowel contrast (e.g., /du̲t/ vs /dy̲t/) at 4 months, but no longer at 10 months of age (Polka & Werker, 1994). Since, many studies with infants from different language backgrounds have provided further evidence that speech sound discrimination towards the end of first year of life improves for many native sound contrasts, accompanied by decreasing discrimination for non-native sounds (overviews in e.g., Byers-Heinlein & Fennell, 2014; Mattock et al., 2008; Maurer & Werker, 2014; Tsuji & Cristia, 2014).

To explain this process of *perceptual attunement* to the native language sound inventory, models of early speech sound perception such as PRIMIR (Processing Rich Information from Multidimensional Interactive Representations, Werker & Curtin, 2005) propose that infants' observable speech discrimination is the result of a dynamic reorganization of speech sound representations shaped by contextual needs and language experience. According to PRIMIR, infants' earliest speech perception is organized on a general perceptual level. Here, phonetic information is encoded following primary auditory processing mechanisms (see also, McCarthy et al., 2019) and modulated by initial biases (e.g., preference for ID-speech, point vowels, a.o.). With growing language experience, these early sound representations are reorganized as a function of the frequency and distribution of features (including acoustic, auditory-visual and articulatory information) encountered in the environment, and the contextual needs (e.g., word learning). Perceptual attunement effects (i.e., increasing native, decreasing non-native discrimination) arise, when infants match incoming instances of vowels based on their similarity with these preliminary experience-driven templates (Werker & Curtin, 2005).

Sound representations continue to undergo important changes as infants build a lexicon (Best et al., 2016; Werker & Tees, 2005). As with non-native contrasts, infants in their late months of the first year show a decline in sensitivity to native sound contrasts that do not differentiate meaning (i.e., allophones, Seidl et al., 2009). Neurally, perceptual attunement parallels a process of 'neural commitment' (Zhang et al., 2005) associated with infants building the sensory-motor schemata for native speech sounds that will allow them to efficiently parse and produce speech and language later on (Kuhl et al., 2005; Kuhl et al., 2014).

Besides language experience, infants' speech sound perception is modulated by a number of factors. First, discrimination ability varies depending on the acoustic properties of sounds themselves. For example, durational differences are discriminated later than spectral differences (e.g., Mugitani et al., 2009; Sato et al., 2010; see also Falk & Tsang, 2020, for a discussion). Discrimination ability is also modulated by linguistic contexts, such as immediately surrounding sounds (Fort et al., 2017; Ko et al., 2009), sound position in the syllable (Archer et al., 2016) or syllabic stress (Floccia et al., 2011). Second, discrimination is influenced by the situational context in which sounds are presented to infants. Infants' discriminate sounds better when they are embedded in multimodal contexts (audiovisual vs. audio or visual only, Ter Schure et al., 2016), and in live interactions vs. videotaped stimuli (Kuhl et al., 2003).

With regard to context, it is largely unknown whether and how infants perceive speech sounds in ID-singing. Indeed, no study to date has addressed this question specifically, and only a few existing studies have attempted to examine 1) whether infants' might at all process verbal information from sung stimuli, and if yes, 2) whether there are developmental differences between speech and song. As to the first question, all four existing studies (François et al., 2017; Lebedeva & Kuhl, 2010; Snijders et al., 2020; Thiessen & Saffran, 2009) suggest that, from very early on, infants extract and process verbal information from stimuli featuring song properties. For example, newborn infants can detect co-occurrences of syllables (i.e., forming pseudo-words) in a syllable stream when every syllable is paired with a different pitch resulting in a melodic contour as in a musical tune (François et al., 2017). However, they fail to do so when syllable streams are presented with flat pitch (François et al., 2017). Older infants (7, 11 months old) detect changes in syllable order when the syllables are presented within musical melodies (Lebedeva & Kuhl, 2010; Thiessen & Saffran, 2009), and extract novel words from a song (Snijders et al., 2020).

As to the second question, however, results are less clear, when comparing speech and song conditions. For word segmentation, Snijders et al. (2020), using ID-like sung children's songs as stimuli, did not find differences between song and speech conditions, and in the study by Lebedeva & Kuhl (2010) on syllable order, a song effect was only close to significance. Thiessen & Saffran (2009) were the only ones to find a difference in speech perception when comparing AD-sung vs. AD-spoken syllables. They familiarized infants aged 6.5-8 months with AD-spoken and AD-sung word sequences (i.e., numbers) which were scrambled in the test phase. Infants listened longer to novel sequences in the sung condition, while no difference between familiar and novel strings were noticeable in the spoken condition. Thiessen and Saffran (2009) interpreted this result as evidence that singing can foster infants' speech processing capacities, possibly because it provides redundant acoustic cues to higher-order structures through consistent mappings between musical and verbal features. For example, many children's songs feature notes that are mapped in a one-by-one manner onto syllables and this mapping forms both higher-order musical (i.e., melodies) and verbal structures (i.e., phrases and utterances). When considering naturalistic ID-singing, acoustic redundancy is even enhanced by the fact that musical features such as pitch, tempo and rhythm are very consistently reproduced by mothers across repeated performances, even more so than in ID-speech (Bergeson & Trehub, 2002). Hence, it is one possibility that consistent co-variation of musical and speech information, as well as consistent reproduction of musical elements over time assist infants' verbal processing in ID-singing and may boost associated learning (e.g., Volkova et al., 2006).

It is another possibility that ID-singing, unlike AD-singing, but similarly to ID-speech, may influence infants' speech processing simply because it attracts infants' attention. A vast amount of literature shows that higher attention to a stimulus typically leads to better processing and learning in infants (e.g., Papageorgiou et al., 2014; Yu et al., 2019). As with ID-speech, infants show more sustained attention to ID-singing than to singing that is not directed at infants (e.g., Trainor, 1996). Infants attend at least as much to ID-singing as to ID-speech (Corbeil et al., 2013; Costa-Giomi & Ilari, 2014), and even prefer to listen to ID-singing compared to ID-speech between 5 and 10 months of age (Nakata & Trehub, 2004; Tsang et al., 2017). Aiming to determine which acoustic features may drive infants' attention to ID-registers, research points towards pitch structure rather than duration or intensity as a particularly powerful acoustic attractor (Fernald & Kuhl, 1987; Räsänen et al., 2018; Thiessen et al., 2005; Thiessen & Saffran, 2009). Expanded pitch contours within vowels and across syllables considerably contribute to infants' higher attending to ID- vs. AD-speech (e.g., Fernald & Kuhl, 1987), and are associated with better outcomes in infants' later attentional and linguistic development (see overview in Spinelli et al., 2017). In ID-singing, flat pitch slopes within vowels forming discrete tonal categories (i.e., notes within a systematic tonal framework) may particularly appeal to infants (Thiessen & Saffran, 2009).

Rhythmic properties such as rhythmic regularity (i.e., temporally predictable occurrences of notes and syllables/words) may as well enhance infants' attending to ID-registers, and particularly to ID-singing. Even foetuses and newborns are sensitive to rhythmic regularity of sounds (Provasi et al., 2014; Winkler et al., 2009), and infants between 5 and 8 month attend longer to regular, structured tone sequences than to irregular ones (Nakata & Mitani, 2005). Although ID-speech may feature more temporally regular word and syllable organisation than AD speech (Malloch, 1999), rhythmic regularity is still more pronounced in ID-singing even when compared to ID-spoken rhymes (Bergeson & Trehub, 2002). Children's songs from the Western tradition mostly display clear metrical structure (e.g., Falk, 2009), that is, the presence of a regular pulse (i.e., beat) and simple recurrent patterns of perceived strong and weak pulse times (i.e., meter: London, 2012). In adults, rhythmic regularity enhances the predictability of speech sounds and thereby, their processing (Cason & Schön, 2012). To date, it remains an open question whether this might apply to infants' processing of speech sounds as well.

In light of the above literature, the present study examined infants' speech sound perception in ID-song compared ID-speech. As stimuli, we chose *vowels* as these carry many of the pitch, duration and spectral features that help identifying ID-registers which infants prefer compared to AD registers (Fernald, 1985; Pegg et al., 1992). Moreover, vowels convey essential information distinguishing speech and singing (e.g., Falk & Tsang, 2020; Sundberg, 1987). For example, the pitch slope within a sung vowel is typically flatter than in speech in order to indicate discrete pitches that form the musical melody within a tonal framework (e.g., Koelsch & Siebel, 2005). Durations of intervocalic intervals play an important role in conveying the rhythmic structure of song (Falk et al., 2014; Sundberg, 1989), whilst inter-syllabic intervals appear to be more relevant for speech (Cutler, 1991). Finally, in some artistic singing styles such as operatic singing, spectral properties of vowels may greatly differ from spoken vowels (e.g., by adding an additional "singer's formant"), due to altered vocal tract constellations (e.g., Sundberg, 1987, 2001).

Generally, vowel intelligibility can be decreased by Western operatic singing style, either because of higher vowel category overlap (Bradley, 2018), or as a result of high-pitched singing (e.g., such as soprano singing, Hollien et al., 2000). In ID-singing, though, which is produced on a continuum with ID-speech (Van Puyvelde & Franco, 2015) and which is deeply rooted in oral lore and song traditions (Mehr et al., 2018), alterations of vowel intelligibility are less likely (e.g., similar to folksong, see Sundberg & Romedahl, 2009). Yet, more corpus research is needed to pinpoint spectral differences of vowels in ID-singing vs. ID-speech (e.g., lesser within-vowel category variability, Audibert & Falk, 2018).

A few studies on ID-speech suggest that ID vowel acoustic characteristics may foster infants' speech sound discrimination. For example, 6-8 and 10-12 months old infants whose mothers display more distinct vowels in ID-speech also show better speech sound discrimination of consonants (Liu et al., 2003). Trainor & Desjardins (2002) tested 6-7 months old infants on their discrimination of the subtle English vowel contrast /I/ (as in *hid*) and /i/ (as in *heed*). They found that the presence of wide pitch range within the intonation contour of a vowel (i.e., similar to saying "Wow!"), a typical prosodic feature of ID-speech (e.g., Fernald & Kuhl, 1987), helps infants to discriminate these fine vowel quality differences. Along similar lines, Adriaans and Swingley (2017) presented evidence from a simulation study that vowels in naturalistic ID-speech, characterized by higher pitch, large pitch range and longer durations, were more likely to enhance infants' phonetic category learning compared to vowels with less extreme values of these acoustic features.

However, not all acoustic ID-features may support infants' vowel sound perception. Trainor & Desjardins (2002) point out that high pitch, as found in both ID-speech and -singing, and flat pitch slope within a vowel, as found specifically in ID-singing, may rather hinder infants' discrimination of fine vowel differences. Several corpus studies also reported higher spectral within-category vowel variability in naturalistic ID- compared to AD-speech, thereby leading to larger overlap of vowel categories (Cristia & Seidl, 2014; Martin et al., 2015). It is thus so far a matter of debate, whether there are sound characteristics of ID-registers that could also negatively influence infants' phonetic learning (Eaves et al., 2016; Martin et al., 2015).

The aim of the present study was twofold. First, given the lack of studies on infants' speech sound perception in ID-singing, we investigated the question of whether infants in their first year of life can discriminate vowels presented in stimuli with typical ID-*singing* acoustics displaying musical melodic and metrical organization. Previous results suggested that these musical features may assist infants' speech processing either because of consistent structural mappings between speech and musical structure or because they generally enhance infants' attending to the stimulus (e.g., Thiessen & Saffran, 2009). Although we will be unable, at this stage, to disentangle these two possibilities, both suggest that infants should indeed show vowel discrimination in ID-singing. The second aim was to examine whether vowel discrimination in ID-singing differs from discrimination in ID-speech. Here, infant age may play a crucial role. Infants younger than 6 months are equally interested in ID-speech and -singing (Corbeil et al., 2013). They are also capable of discriminating a wide range of native and non-native language speech sounds (e.g., Chladkova & Paillereau, 2019, Tsuji & Cristia, 2014, for overviews). Provided that both types of ID-acoustics are adequate to help infants' vowel discrimination, we expect that this group of infants may show similar patterns of vowel discrimination when listening to either ID-speech or ID-singing.

However, in the second half year, as infants become attuned to language- as well as music-specific perception (e.g. Soley & Hannon, 2010), discrimination may diverge. On the one hand, language-specific sound processing may be more relevant to infants' ID-speech perception as ID-speech interactions in their daily lives become guided by joint attentional exchanges and increasingly centred around linguistic functions of sound contrasts (e.g., such as distinguishing words in referential contexts, e.g., Fennell & Waxman, 2010). On the other hand, increasing musical proficiency may heighten older infants' processing abilities of musical structure resulting in higher attending towards ID-singing (as found in Tsang et al., 2017). Hence, the effects of perceptual attunement (e.g., decrease in non-native vowel discrimination) may be more visible with ID-spoken than ID-sung stimuli. In sum, infants' perception in the second half year of life may be increasingly affected by linguistic and / or musical experience and associated functions, and thereby bias infants' vowel discrimination, and the perceptual attunement effects we observe in ID-speech and -singing.

In two experiments, using a preferential listening task (Best & Jones, 1998), we presented ID-like spoken and sung stimuli featuring typical ID-acoustics (e.g., speech contour with expanded pitch range, sung contour featuring a clear melodic and metrical pattern) to Italian-learning infants in their first and second half year of life, i.e., before and after the start of language-specific vowel perception (i.e., around 6 months, see Tsuji & Cristia, 2014). The stimuli contained a novel non-native (German) vowel contrast which was chosen to allow us to observe potential perceptual attunement effects in younger and older infants, and to exclude any familiarity effect. For infants under 6 months of age (Exp.1), we predict discrimination of the novel vowel in both ID-like singing and speech, along the lines of well-established previous research on infants' early sound perception in speech (e.g., Tsuji & Cristia, 2014; Werker & Tees, 2005). Such a result would suggest that acoustic ID-characteristics, whether sung or spoken are well-suited to sustain infants' attention and associated early auditory perception skills. For infants in their second half year of life (Exp. 2), however, the perceptual attunement literature suggests that the discrimination of the non-native vowel contrast will deteriorate (Kuhl et al., 2008; Maurer & Werker, 2014). We examine if this will be indeed the case in both ID-speech and -singing, or whether, based on potential effects of musical / linguistic experience on infants' perception, older infants will continue to discriminate the contrast in ID-singing, but not ID-speech.

Experiment 1

Method

Participants

As there were no previous studies available on infants' speech sound perception in speech vs. singing using naturalistic stimuli, we ran a Power analysis after having collected data from a small sample to decide on final sample size. Data from 14 infants were collected who listened either to the familiar and novel vowels presented in the ID-sung stimuli (n = 7), or presented in the ID-speech stimuli (n = 7). In order to obtain an estimation of effect size, a mixed factorial Analysis of Variance (ANOVA) was run with the between-subject factor Register (i.e., ID-speech/singing) and Trial (familiar vs. novel vowel) as a within-subject factor. We found a large effect size, $\eta_p^2 = 0.387$, for the hypothesized within-subject effect of Trial (i.e., discrimination between familiar and novel vowels) which was entered into G*Power 3.1.9.4 (Faul et al., 2007) with alpha = .05. Suggested overall sample size was n = 22, with a Power of .92.

Hence, our final sample consisted of 24 infants, randomly assigned to the speech or singing condition. Five additional infants were tested, but excluded from the analyses due to fussiness. All participants were healthy full-term infants from monolingual Italian households aged between 4 and 6 months (M = 4.7; SD = 0.64). Infants lived in predominantly white middle-class households in Milano, Northern Italy. Infants in the two conditions were similar in age, sex, maternal education, and on their familiarity with singing at home (see Table 1). The latter was assessed by caregivers' answers to the Italian adaptation of the Music@Home questionnaire (Politimou et al., 2018) on family musical interactions at home. The *Exposure to singing scale* score from the questionnaire was used to match infants' in both groups according to familiarity with ID-singing in their everyday life. Based on both this score (p = 0.65) and maternal education (p = 0.2), the experimental groups were equivalent.

Ethics

The present study was conducted according to guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from a parent for each child before any assessment or data collection. All procedures involving human subjects in this study were approved by the Psychology Research Ethics Committee at Milan-Bicocca University. Parents received an information sheet, signed written consent and were provided with debriefing at the end of their participation.

Stimuli

To test phonetic discrimination with Italian-learning infants, the German vowel contrast /u/ - /y/ was chosen, a contrast that has been previously studied in English-learning infants (see overview in Polka & Bohn, 2011). Figure 1 displays the quality of the German /u/ and /y/ vowels used in the experiment in relation to vowels in the Italian variety of the region where parents were recruited (Savy & Cutugno, 1997). The vowel /u/, such as in the English word "shoe", is a native-like vowel that resembles the quality of Italian /u/ (see Fig. 1). In Italian, at least adult speakers easily assimilate this vowel into the native vowel category (Missaglia, 2004). Differently, /y/ (pronounced in English as "e" with rounded lips, as in the German city name "M**ü**nchen") is a novel sound for Italian-learning infants as it is not part nor resembles the standard or regional Italian vowel system (Krämer, 2009; Savy & Cutugno, 1997). As displayed in Fig. 1, formant values of /y/ are outside the range of other Italian vowel categories.

Vowels were embedded in simple open syllables starting with the plosive /g/ (i.e., "gu" or "gü") to enhance sonority contrasts (i.e., the vowel was clearly more sonorous than the consonant). Unlike previous studies on phonetic discrimination which mainly used single isolated syllables and monotone melodic and rhythmic contours, we embedded the vowel contrast in a contour modelled after ecological ID-singing and -speech. The ID-like singing stimuli were musically more elaborate than the ID-like speech stimuli. Melodic and rhythmic differences are displayed in Fig. 2.

In both the ID-like speech and the ID-like singing versions, the syllable sequences formed a smoothly rising and then falling (i.e., bell-shaped) melodic contour which parents use when they communicate approval or appraisal of their child's behavior (Papoušek et al., 1991, see Fig. 2). In the ID-like singing version, syllable sequences featured clear melodic structure implemented as systematic recurrences of stable pitch relations (i.e., musical intervals) constrained by a tonal framework (i.e., the western tonal system based on scales of 12 semitones). Each syllable had a discrete pitch value corresponding to a note in order to create a three-syllable ascending-descending melodic contour progressing in intervals of thirds and spanning the range of a perfect fifth (D4-F#4-A4-A4-F#4-D4). In the ID-like speech version, the pitch contour (with the peak on the third syllables, as in ID-like singing) was achieved by dynamic pitch variation inside each syllable and an overall extended pitch range as in naturalistic ID-speech (Fernald & Kuhl, 1987). ID-like sung versions also featured a clear musical rhythm (a triple meter), implemented as recurring triples of one strong (i.e., longer) and two weak (i.e., shorter) syllables (imagine steps in a waltz, ONE–two-three). As a consequence, one /u/ and one /y/ vowel were associated to these prominent positions in song. ID-like speech stimuli did not have a clear metric structure. Note that these rhythmic manipulations induced a significant difference in durational variability of vowels (i.e., ID-like song > ID-like speech, see Fig. 2). Both types of stimuli displayed similar final lengthening on the last syllable of the contour (see Fig. 2).

A female native speaker of German was recorded (at 44 kHz sampling rate, 16 bit) in a sound-proof booth while singing and speaking sequences of six "gu" and "gü" syllables. Uniform sequences exclusively contained the native-like vowel (6 sung or spoken /gu/-syllables). Varied sequences contrasted the non-native with the native-like vowel (6 sung or spoken syllables in the order /gy-gu-gy-gu-gy-gu/, see Fig. 2). The speaker was instructed to sing and speak both types of sequences as if an infant was present, in an engaging ID-like manner, but without overt smiling in order to preserve vowel qualities. We chose one uniform and one varied syllable sequence (2.45 s long) from the recordings per speech and singing condition to construct the trial structure of the experiment (see Mattock et al., 2008). All sequences were normalized in amplitude. Uniform sequences (i.e., exclusively /gu-gu-gu-gu-gu-gu/) were repeated 10 times with a pause of 0.5 sec between sequences to form a non-alternating trial (~ 30s total duration). An alternating trial consisted of the alternation of the varied with the uniform sequence (in the order /gy-gu-gy-gu-gy-gu/ followed by /gu-gu-gu-gu-gu-gu/) until 30s with, again, a 0.5s pause between sequences.

Acoustic variables (measured with PRAAT, Boersma & Weenink, 2019) of ID-like speech and singing are displayed in Table 2. ID-like-speech and -singing stimuli displayed typical pitch range differences observed in naturalistic interactions between parents and babies. That is, overall pitch range was smaller in sung than in spoken stimuli (e.g., Tsang et al., 2017). Mean inter-syllable interval and mean vowel duration were comparable in ID-like speech and singing. Thus, we made sure that, in the test phase, infants were exposed to each vowel for an equal amount of time in both conditions. As noted above, vowel duration variability (as evidenced by higher SD, Table 2) was considerably higher in ID-like singing compared to speech, due to the strong metric structure of song. Vowel quality was determined by measuring the first two formants (F1, F2; see Table 2 for a summary) at the midpoint of each /u/ and /y/

vowel (using the Burgh algorithm combined with visual inspection in PRAAT). Quality of /u/ (i.e., F1 and F2) in uniform (F1: *M* = 394 Hz, *SD*= 22 Hz; F2: M= 922 Hz, *SD*= 57 Hz) vs. varied sequences (F1: *M* = 404 Hz, *SD*= 18 Hz; F2: *M*= 968 Hz, *SD*= 35 Hz) in either condition did not significantly differ nor did quality of /u/ in speech vs. singing (see Fig. 1 and mean values in Table 2). Vowel quality of /y/ was clearly different from /u/ (i.e., fronted articulation of /y/, as indicated by a clearly higher second formant F2, see Fig. 1, and Table 2) in both spoken and sung stimuli (Mann-Whitney, *Z*= 2.50, *p* = .009, in both conditions).

*Procedur*e

Vowel discrimination was tested using the Preferential Listening with Alternating Trial Procedure (Best & Jones, 1998), modeled on the protocol described by Mattock et al. (2008). In the standard Alternating Trial Procedure, infants are familiarized with a single syllable / sound. Sound discrimination is tested by presenting a single syllable that contains a sound (non-alternating trial) on half of the trials, and in the remaining trials presenting two syllables in alternation (alternating trial), one of which contains a novel sound. Infants' longer looking times to one of the trial types, typically the alternating trial, are interpreted as discrimination of the sound contrast. Our novel manipulation consisted in presenting the syllables as sequences grouped under an ID-like spoken or sung contour (see stimuli and Fig. 2).

Familiarization consisted of non-alternating sequences presenting only /gu/ syllables with the native-like /u/ vowel sound in either the ID-like speech or the singing condition, until 30 cumulative seconds of listening time were accrued. Note that the order of /u/ in the familiarization with a subsequent introduction of /y/ in the Test phase should make the task more difficult for infants unfamiliar with German than the reverse order (e.g., Polka & Bohn, 2011). After Familiarization, the Test phase immediately started. Infants listened to four alternating and four non-alternating trials (e.g., Alt, Non-Alt, Alt, Non-Alt, and so on). The trial type with which infants started the Test phase (alternating vs. non-alternating) was counterbalanced across participants.

The experiment was run using E-Prime 2.0-SP2 software in an anechoic chamber. Caregivers were wearing headphones playing sound-masking white noise throughout the experiment while their infants were sitting on their lap. Infants faced a computer screen displaying a colorful geometrical pattern. Loudspeakers were localized behind the screen. The experimenter monitored infant gaze from outside the chamber, via a muted video-link with a close-up camera and entered information about infant looking behavior by button press. Infant looking time towards the screen was recorded as *listening time* for each trial. When the infant looked away for more than 2 seconds, the trial was interrupted, and an attention getter (i.e., a bright red expanding circle) appeared in the center of the screen; when the infant looked back towards the screen, the next trial started.

Results

Statistical analyses were conducted using a 2 (*Register*: ID-like speech/singing) X 2 (*Trial*: familiar/novel vowel) mixed factorial Analysis of Variance (ANOVA) on listening time per trial. As shown in Fig. 3, infants were able to discriminate the sound contrast. They were listening considerably longer to the alternating trials containing the novel, non-native vowel than to the non-alternating trials, with a large effect of Trial, $F(1,22) = 11.48$, $p = .003$, $\eta_p^2 = 0.34$. Neither the effect of Register, $F(1,22) = 0.149$, $p = .7$, $\eta_p^2 = 0.007$, nor the Register x Trial interaction, $F(1,22) = 0.654$, $p = .43$, $\eta_p^2 = 0.029$, became significant, that is, discrimination rates in the ID-like speech and -singing conditions were not statistically different. Overall, 18 (7 in the ID-like speech, 11 in ID-like singing condition) out of 24 infants showed a strong preference for novelty (i.e., greater listening time during alternating trials).

Experiment 2

Using the same design as in Experiment 1, Experiment 2 aimed to assess whether infants at 8-10 months of age were still able to discriminate the vowel contrast used in Experiment 1.

Method

Participants

Twenty-four healthy and full-term 8 to 10 months old infants (*M* = 8.4 months, *SD* = 0.65) from monolingual Italian, predominantly white middle-class households from the Milano region (Northern Italy) were randomly assigned to a Speech or Singing condition. The groups were similar on key background variables including *Exposure to singing* (p = 0.68) and maternal education (p = 0.75; see Table 3). Again, participant recruitment was guided by an estimated sample size based on a Power analysis (ANOVA, 2 x 2 mixed factorial design) from data of n=7 participants in each Register (i.e. ID-like speech/singing). Based on this preliminary data, we found a large effect size, $\eta_p^2 = 0.347$, for the crucial within-between subject interaction Trial x Register. The results provided by G*Power (with alpha = .05) suggested an overall sample size of n=24, with a Power of .90.

Stimuli & Procedure, Ethics

The same stimuli, experimental setting, and ethics procedure were used as in Experiment 1.

Results

Statistical analyses were conducted using a 2 (*Register*: ID-like speech/singing) X 2 (*Trial:* familiar / novel vowel) mixed factorial ANOVA on listening time during the trials. The analysis revealed a significant moderate *Register X Trial* interaction, $F(1, 22) = 4.38$, $p = .048$, $\eta_p^2 = 0.166$ (see Fig. 4), with non-significant main effects of both *Trial,* $F(1, 22) = 3.44$, $p = .077$, $\eta_p^2 = 0.135$, and *Register,* $F(1, 22) = 0.002$, $p = 0.96$, $\eta_p^2 < 0.001$. Decomposing the interaction using planned t-tests showed that only infants listening to ID-like singing discriminated between the novel and familiar vowel (ID-like singing: $t(11) = -2.81$, $p = .017$, $d = 0.82$; ID-like speech: $t(11) = .17$, $p = .87$, $d = 0.049$). As in Experiment 1, the majority of infants (15 out of 24; 6 in the speech, 9 in the singing condition) showed a preference for novelty (i.e., greater listening time during alternating trials).

Discussion

The first aim of the present study was to test whether infants in their first year of life can discriminate speech sounds in stimuli featuring properties of ID-singing, a musical input which plays an important role in early parent-infant interactions (Falk & Tsang, 2020; Malloch & Trevarthen, 2009; Trehub & Gudmondsdottir, 2015). The second aim was to test whether infants' sound discrimination in sung stimuli would differ from stimuli featuring ID-speech acoustics in the first or second half year of life. We presented a vowel contrast including a familiar, native-like vs. a novel, non-native vowel in stimuli modeled after the melodic and rhythmical structure of ecological ID-singing and ID-speech. Infants of two age groups listened to the stimuli, one group in the first half of their first year of life, and one group in the second half, in order to test potential effects of perceptual attunement (i.e., loss of non-native vowel discrimination in the second half of the first year of life, see Tsuji & Cristia, 2014) on infants' sound perception in both conditions.

As to our first aim, results revealed that infants in both age groups discriminated vowels in ID-like singing. This finding adds to recent evidence (François et al., 2017; Lebedeva & Kuhl, 2010; Snijders et al., 2020; Thiessen & Saffran, 2009) that infants do process speech information, and more specifically *speech sound* information, in singing. As to the second aim, we found differences in vowel discrimination between ID-like speech and singing depending on age group. Consistently with the literature on vowel discrimination in speech (e.g., Tsuji & Cristia, 2014), infants in the first half year of life discriminated the novel from the familiar vowel in both ID-like singing and ID-like speech. This result indicates that the acoustics of both ID-like sung and spoken stimuli were appropriate for infants' vowel sound processing. Although some of the features were present that were previously found to decrease vowel discrimination (Trainor & Desjardins, 2002), such as high pitch (particularly present in our spoken stimuli) and discrete pitches (particularly present in our sung stimuli), infants in the younger age group successfully discriminated the vowel contrast in both types of stimuli.

Older infants in their second half year of life, however, who typically are attuning to native sounds (e.g., Kuhl et al., 2008), only discriminated the novel vowel in the sung, but not in the spoken condition. For ID-like speech, this finding replicates the well-established developmental pattern of non-native sound discrimination (Kuhl et al., 2008; Werker & Tees, 1984, 2005, amongst others). However, it does so with a new type of stimulus that was acoustically more elaborate (i.e., possessing an intonational and rhythmic contour) compared to monotone, isolated syllables typically used in similar tests on infants' speech sound discrimination (e.g., Kuhl et al., 2008). This finding may be useful to pave the way to new studies employing more naturalistic stimuli, thus enhancing the ecological validity of infant discrimination experiments. Importantly, the finding for ID-like singing reveals that infants can discriminate novel speech sounds in sung stimuli even past the typical time window for non-native sound discrimination in speech.

These results open interesting questions about the contribution of musical characteristics to speech sound discrimination, and about potential processing differences between ID-speech and ID-singing. First, we will address the question of which of the musical ID-acoustics in the sung stimuli could have influenced the pattern of discrimination found in both age groups. Second, potential underlying mechanisms of infants' vowel sound perception in ID-singing vs. ID-speech will be discussed.

Following the argument of Thiessen and Saffran (2009), some musical features contained in our ID-like sung stimuli may have made the differences between vowels more available to infants' discrimination, or at least have enhanced attending towards sung stimuli. Similarly to many traditional and modern songs for babies and children, our ID-like sung stimuli featured discrete pitch-syllable associations (1 syllable – 1 note) forming a simple memorable melody. These one-to-one mappings which enhance redundancy of acoustic cues to both verbal and musical structure (Thiessen & Saffran, 2009), have been reported to appeal to infants and to enhance their processing of verbal information (see François et al., 2017). Although our stimuli were composed of repeated syllables instead of words, it is a possibility that the consistent mapping between melodic and syllabic structure has influenced infants' perception of syllables and vowel sounds therein in both age groups.

Increasing musical experience may have facilitated older infants' attending and processing of musical features in the ID-like sung stimuli. As a result, the musical interval structure and some rhythmic characteristics of our ID-like sung stimuli may have elicited the interest of the older infants, in particular. As they become more proficient in processing musical melodies, musical interval structure is more appealing to infants in their second half than in their first half year of life (Schellenberg & Trehub, 1996; Trainor & Trehub, 1992; Trehub et al., 1984).

Concerning rhythmic features, our ID-like sung stimuli displayed a clear metrical pattern (a triple meter, as in a waltz rhythm) compared to the ID-like spoken stimuli (no specific metrical pattern). Although infants can discriminate melodies with different metrical properties from 4 months on (e.g., Soley & Hannon, 2010), only in their second half year of life do they use metrical patterns in music and speech to efficiently process these sound sequences. For example, from 7.5 months on, infants start relying on metrical syllable patterns to segment their first words from continuous speech (e.g., Jusczyk et al., 1999). Stronger metrical structure in musical sequences also enhances older infants' perception of fine temporal and melodic changes (Bergeson & Trehub, 2006; Trehub & Hannon, 2009). Future studies could further explore whether this perceptual advantage pertains to spectral changes in speech sounds associated with a musical sequence.

In addition to these musical features, there are some other features that could have contributed to infants' vowel perception in ID-like singing. One novel vowel was presented in a metrically strong position, which could have made this particular vowel more prominent and easier to process for infants in ID-like singing compared to ID-like speech. Durations of vowel sounds were also more variable in ID-like singing compared to ID-like speech, which could potentially have attracted more attention. Note that both properties (metrical strength, higher durational variability of syllables) are a consequence of the naturalistic metrical structure that is typical of children's playsongs (Trehub & Trainor, 1998). However, there were also several naturalistic ID-speech characteristics in our stimuli such as higher pitch and larger pitch range that infants usually prefer, even more so than durational aspects (Fernald & Kuhl, 1987; Thiessen et al., 2005). As the present study was not designed to test the contribution of individual acoustic features, and in the absence of a main effect of register in the experiments, it remains an open question whether infants attended more to one specific property of the stimuli. Future research will determine which exact combination of acoustic ID-characteristics (including prominence, variability, pitch structure) most impact infants' speech sound perception in complex sung and spoken auditory stimuli.

Besides the account that infants' speech sound discrimination in ID-singing is influenced by musical features, an alternative suggestion is that infants *process* speech sounds in singing differently than in speech. This idea is in line with context-dependent models of infants' speech perception, such as the framework for Processing Rich Information from Multidimensional Interactive Representations (PRIMIR, Werker & Curtin, 2005) or the Parallel Channels Model (PCM, Gallé & McMurray, 2014). These models postulate that infants use different processing modes depending on their contextual needs, which may change, for example, when infants start learning words in their native language (Stager & Werker, 1997;

Werker & Curtin, 2005). The PCM model, in addition, proposes that, during development, infants can use both an acoustically grounded perception mode and categorical perception in parallel (Gallé & McMurray, 2014, based on Pisoni & Tash, 1974). Infants would co-improve both modes constantly with age, but younger infants would rely more on the acoustic channel than older infants (see McCarthy et al., 2019, for spectral vowel properties, and the discussion in Mugitani et al., 2009, for vowel length properties).

Accordingly, the younger infants in our study may have been driven by an acoustic approach to discriminate acoustical differences of vowels in both ID-like speech and singing (e.g., relying on the acoustic energy structure of the vowels, see the NRV framework of Polka & Bohn, 2011). Instead, the older infants - more heavily influenced by language experience - may have started to rely more on language-specific vowel organisation in speech and may have assimilated both vowels into the same native category for /u/ (see e.g., Kuhl et al., 2008). However, in sung stimuli, they may have continued to discriminate the vowels similarly to the younger age group, based on acoustic information and, potentially, initial auditory perceptual biases (e.g., Polka & Bohn, 2011).

Such potential differences in processing modes could originate from differences in familiarity and daily exposure to singing and speech (Bergelson et al., 2019; Trehub & Trainor, 1998) or from the development of different cognitive mechanisms and neural activations underlying language and music processing (Angulo-Perkins & Concha, 2019; Best et al., 1982; Dehaene-Lambertz et al., 2010). Finally, it is also possible that sung vowels, because of their musical acoustics, are not perceived as a good fit to the phonological categories infants are developing for speech, and therefore are processed differently from spoken vowels. These are interesting issues to address in future studies aiming to disentangle these possibilities.

The present results reveal a longer discrimination window for older infants' *non-native* sound perception in ID-like singing compared to ID-like speech. This is consistent with the finding that infants of this age are particularly interested in ID-singing in a foreign language

(Tsang et al., 2017). It is unknown whether non-native sound perception is simply protracted in ID-singing, or whether non-native sound discrimination will persist into later infancy or even childhood. However, there is a suggestion that the latter hypothesis might be correct. Schön et al. (2008) found that some aspects of speech perception in an artificial novel language were facilitated in a sung context for both children and adults, compared to a spoken context. The authors suggested that musical structure may play a facilitating role particularly *at the very early stages* of acquiring a new language. Along similar lines, Lebedeva and Kuhl (2010) concluded from their own study and other results obtained with older children (Calvert & Billingsley, 1998), that there might be a phonetic advantage when encoding words from song which could start developing as early as the first year of life.

It would also be highly relevant to test whether our finding for non-native sound discrimination in singing generalizes to *native* vowel discrimination. Better native sound discrimination in speech is associated with infants' verbal production and comprehension abilities 2 years later (Kuhl et al., 2005; Tsao et al., 2004). Therefore, future studies could assess whether discrimination of native language sounds is equally enhanced in singing and whether this is predictive of infants' later language abilities. In addition, native vs. non-native sound perception could be tested with different populations such as monolingual and bilingual infants. Infants growing up or transitioning to a bilingual environment (e.g., in minority contexts or migration) could particularly benefit from a longer time window for non-native sound discrimination in ID-singing when learning novel or less familiar speech sounds from song.

One limitation of our study was that our stimuli were only modeled on the acoustics of ecological stimuli (i.e., ID-*like*-speech and -singing). Thus, some acoustic differences between naturalistic ID-speech and -singing were not present, such as generally longer vowels in singing than speech (Tsang et al., 2017), or spectral differences in vowels caused by, for instance, more smiling during ID-singing than ID-speech (Trehub et al., 2016). As infants may prefer these features, at least at some period during the first year (e.g., Kitamura & Notley, 2009; Singh et al., 2002), naturalistic ID-singing could even be more attractive to them, which in turn could impact vowel perception. Our stimuli also lacked the rich social, multisensory and emotional experiences of live interaction, which enhance infants' learning. For example, 9-month-olds exposed to live interaction with adults are more able to learn non-native phonemes than infants only exposed to audio-visual recordings (Kuhl et al., 2003). Active musical participation of infants was also found to foster social and preverbal communicative skills compared to passive musical experiences (Gerry et al., 2012). Hence, it is likely that ID-singing set in a social context of engaging interaction would have a higher impact on infants' abilities to perceive and potentially learn new sounds from song.

The present study is a first step towards understanding the underlying mechanisms of speech perception in ID-singing. More extensive testing with larger sample sizes, different sounds, and different populations (e.g., older children, bilingual learners etc.) will shed light on how infants perceive and potentially learn speech from song. For example, a future training study could test long-term effects on infants' phonological and phonetic learning from song and the role of musical features therein. As recently shown in an infant training study, active music lessons (including movement to songs) can improve infants' temporal perception in both music and speech (Zhao & Kuhl, 2016).

Finally, our findings enrich a growing body of research showing that music matters for infants' development. Musical stimulation can enhance infants' physiological well-being (Van der Heijden et al., 2016), facilitate prosocial behaviour, as well as arousal and affect regulation (e.g., Cirelli et al., 2014; Mehr & Spelke, 2018; Trehub, Ghazban et al., 2015). In light of these studies, ID-singing may be of particular relevance to early education and intervention programs. One advantage of using ID-singing in such programs is its informal nature and its availability in many spontaneous caregiving contexts without preparation or instrumental support. Moreover, first results in this domain show that exposure to caregivers' ID-singing may contribute to generally strengthening the interweaving of music and language development with possible benefits for children's first and second language learning (Forgeard et al., 2008; Politimou et al., 2019). In conclusion, the results of this study open the door to a new generation of research and future applications into the role of singing with infants to assist language development.

References

Adriaans, F., & Swingley, D. (2017). Prosodic exaggeration within infant-directed speech: Consequences for vowel learnability. *The Journal of the Acoustical Society of America, 141*(5), 3070. https://doi.org/10.1121/1.4982246

Albin, D. D., & Echols, C. H. (1996). Stressed and word-final syllables in infant directed speech. *Infant Behavior and Development, 19*, 401-418. http://dx.doi.org/10.1016/S0163-6383(96)90002-8

Angulo-Perkins, A., & Concha, L. (2019). Discerning the functional networks behind processing of music and speech through human vocalizations. *PloS One*, *14*(10), e0222796. https://doi.org/10.1371/journal.pone.0222796

Archer, S. L., Zamuner, T., Engel, K., Fais, L., & Curtin, S. (2016). Infants' discrimination of consonants: Interplay between word position and acoustic saliency. *Language Learning and Development, 12*(1), 60-78. https://doi.org/10.1080/15475441.2014.979490

Audibert, N., & Falk, S. (2018). Vowel space and f0 characteristics of infant-directed singing and speech. *Proceedings of the 9th International Conference on Speech Prosody 2018*, 153-157. https://doi.org/10.21437/SpeechProsody.2018-31

Bergelson, E., Casillas, M., Soderstrom, M., Seidl, A., Warlaumont, A. S., & Amatuni, A. (2019). What do North American babies hear? A large-scale cross-corpus analysis. *Developmental Science, 22*:e12724. https://doi.org/10.1111/desc.12724

Bergeson, T. R., & Trehub, S. E. (2002). Absolute pitch and tempo in mothers' songs to infants. *Psychological Science*, *13*(1), 72-75. https://doi.org/10.1111/1467-9280.00413

Bergeson, T. R., & Trehub, S. E. (2006). Infants perception of rhythmic patterns. *Music Perception, 23*(4), 345-360. https://doi.org/10.1525/mp.2006.23.4.345

Bernstein-Ratner, N. (1986). Durational cues which mark clause boundaries in mother-child speech. *Journal of Phonetics, 14*, 303–309. https://doi.org/10.1016/S0095-4470(19)30670-9

Best, C.T., & Jones, C. (1998). Stimulus-alternation preference procedure to test infant speech discrimination. *Infant Behavior and Development*, *21*, 295. https://doi.org/10.1016/S0163-6383(98)91508-9

Best, C.T., Hoffman, H. & Glanville, B.B. (1982). Development of infant ear asymmetries for speech and music. *Perception & Psychophysics, 31*, 75-85. https://doi.org/10.3758/BF03206203

Boersma, P., & Weenink, D. (2019). Praat: doing phonetics by computer [Computer program]. Version 6.0.50, retrieved 31 March 2019 from http://www.praat.org/

Bombar, M. L., & Littig, L. W., Jr. (1996). Babytalk as a communication of intimate attachment: An initial study in adult romances and friendships. *Personal Relationships, 3*(2), 137-158. https://doi.org/10.1111/j.1475-6811.1996.tb00108.x

Bornstein, M. H., Hahn, C. S., Suwalsky, J., & Haynes, O. M. (2011). Maternal and infant behavior and context associations with mutual emotion availability. *Infant Mental Health Journal*, *32*(1), 70–94. https://doi.org/10.1002/imhj.20284

Bradley, E.D. (2018). A comparison of the acoustic vowel spaces of speech and song. *Linguistic Research 35*(2), 381-394. https://doi.org/10.17250/khisli.35.2.201806.006

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science, 296*(5572), 1435. https://doi.org/10.1126/science.1069587

Byers-Heinlein, K., & Fennell, C. T. (2014). Perceptual narrowing in the context of increased variation: Insights from bilingual infants. *Developmental Psychobiology*, *56*(2), 274-291. https://doi.org/10.1002/dev.21167

Calvert, S. L., & Billingsley, R. L. (1998). Young children's recitation and comprehension of information presented by songs. *Journal of Applied Developmental Psychology, 19*(1), 97–108. https://doi.org/10.1016/S0193-3973(99)80030-6

Cason, N., & Schön, D. (2012). Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia*, *50*(11), 2652-2658. https://doi.org/10.1016/j.neuropsychologia.2012.07.018

Chládková, K., & Paillereau, N. (2020). The What and When of universal perception: A review of early speech sound acquisition. *Language Learning, 70*, 1136-1182. https://doi.org/10.1111/lang.12422

Cirelli, L. K., & Trehub, S. E. (2020). Familiar songs reduce infant distress. *Developmental Psychology*, *56*(5), 861. https://doi.org/10.1037/dev0000917

Cirelli, L. K., Einarson, K. M., & Trainor, L. J. (2014). Interpersonal synchrony increases prosocial behavior in infants. *Developmental Science*, *17*(6), 1003-1011. https://doi.org/10.1111/desc.12193

Cirelli, L. K., Jurewicz, Z. B., & Trehub, S. E. (2020). Effects of maternal singing style on mother–infant arousal and behavior. *Journal of Cognitive Neuroscience*, *32*(7), 1213-1220. https://doi.org/10.1162/jocn_a_01402

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development, 61*(5), 1584-1595. https://doi.org/10.2307/1130766

Corbeil, M., Trehub, S. E., & Peretz, I. (2013). Speech vs. singing: Infants choose happier sounds. *Frontiers in Psychology, 4*, 1-11. https://doi.org/10.3389/fpsyg.2013.00372

Corbeil, M., Trehub, S. E., & Peretz, I. (2016). Singing delays the onset of infant distress. *Infancy*, *21*(3), 373-391. https://doi.org/10.1111/infa.12114

Costa-Giomi, E., & Ilari, B. (2014). Infants' preferential attention to sung and spoken stimuli. *Journal of Research in Music Education, 62*, 188-194. https://doi.org/10:1177/0022429414530564

Cristia, A., & Seidl, A. (2014). The hyperarticulation hypothesis of infant-directed speech. *Journal of Child Language*, *41*(4), 913-934. https://doi.org/10.1017/S0305000912000669

Cutler, A. (1991). Linguistic rhythm and speech segmentation. In J. Sundberg, J. N. Lennart, & R. Carlson (Eds.), *Music, language, speech and brain* (pp. 157-166). Macmillan

Dehaene-Lambertz, G., Montavont, A., Jobert, A., Allirol, L., Dubois, J., Hertz-Pannier, L., & Dehaene, S. (2010). Language or music, mother or Mozart? Structural and environmental influences on infants' language networks. *Brain and Language*, *114*(2), 53-65. https://doi.org/10.1016/j.bandl.2009.09.003

Eaves, B. S., Feldman, N. H., Griffiths, T. L., & Shafto, P. (2016). Infant-directed speech is consistent with teaching. *Psychological Review*, *123*(6), 758-771. https://doi.org/10.1037/rev0000031

Falk, S. & Tsang, C.D. (2020). The role and functions of infant-directed singing in early development. In F. Russo, B. Ilari, & A. Cohen (Eds.), *The Routledge Companion to Interdisciplinary Studies in Singing, Volume I: Development* (pp.179-188). Routledge. https://doi.org/10.4324/9781315163734

Falk, S. (2009). *Musik und Sprachprosodie. (Music and Speech prosody)*. De Gruyter.

Falk, S. (2011). Melodic versus intonational coding of communicative functions: A comparison of tonal contours in infant-directed song and speech. *Psychomusicology: Music, Mind and Brain*, *21*(1-2), 54-68. https://doi.org/10.1037/h0094004

Falk, S., Rathcke, T., & Dalla Bella, S. (2014). When speech sounds like music. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1491-1506. https://doi.org/10.1037/a0036858

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191. https://doi.org/10.3758/BF03193146

Fennell, C. T., & Waxman, S. R. (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child development*, *81*(5), 1376-1383. https://doi.org/10.1111/j.1467-8624.2010.01479.x

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior & Development, 8*(2), 181-195. https://doi.org/10.1016/S0163-6383(85)80005-9

Fernald, A. (1989). Intonation and communicative content in mothers' speech to infants: Is the melody the message? *Child Development, 60*, 1497-1510. https://doi.org/10.2307/1130938

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development, 10*(3), 279-293. https://doi.org/10.1016/0163-6383(87)90017-8

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology, 20*(1), 104-113. http://dx.doi.org/ 10.1037/0012-1649.20.1.104

Fernald, A., Taeschner, T., Dunn, J., Papoušek, M., Boysson-Bardies, B. de, & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language, 16*(3), 477-501. https://doi.org/10.1017/S0305000900010679

Filippa, M., Devouche, E., Arioni, C., Imberty, M., & Gratier, M. (2013). Live maternal speech and singing have beneficial effects on hospitalized preterm infants. *Acta paediatrica*, *102*(10), 1017-1020. https://doi.org/10.1111/apa.12356

Floccia, C., Nazzi, T., Austin, K., Arreckx, F., & Goslin, J. (2011). Lexical stress and phonetic processing in word learning in 20- to 24-month-old English-learning children. *Developmental Science, 14*(3), 602-613. https://doi.org/10.1111/j.1467-7687.2010.01006.x

Forgeard, M., Winner, E., Norton, A., & Schlaug, G. (2008). Practicing a musical instrument in childhood is associated with enhanced verbal ability and nonverbal reasoning. *PloS One*, *3*(10), e3566. https://doi.org/10.1371/journal.pone.0003566

Fort, M., Brusini, P., Carbajal, M. J., Sun, Y., & Peperkamp, S. (2017). A novel form of perceptual attunement: Context-dependent perception of a native contrast in 14-month-old infants. *Developmental Cognitive Neuroscience, 26*, 45-51. https://doi.org/10.1016/j.dcn.2017.04.006

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), 12451. https://doi.org/10.1038/s41598-017-12798-2

Gallé, M. E., & McMurray, B. (2014). The development of voicing categories: A quantitative review of over 40 years of infant speech perception research. *Psychonomic Bulletin & Review, 21*(4), 884-906. https://doi.org/10.3758/s13423-013-0569-y

Gerry, D., Unrau, A., & Trainor, L. J. (2012). Active music classes in infancy enhance musical, communicative and social development. *Developmental Science*, *15*(3), 398-407. https://doi.org/10.1111/j.1467-7687.2012.01142.x

Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., & Parlato-Oliveira, E. (2015). Early development of turn-taking in vocal interaction between mothers and infants. *Frontiers in Psychology, 6*, 1167. https://doi.org/10.3389/fpsyg.2015.01167

Hartman, K. M., Ratner, N. B., & Newman, R. S. (2017). Infant-directed speech (IDS) vowel clarity and child language outcomes. *Journal of Child Language*, *44*(5), 1140-1162. https://doi.org/10.1017/s0305000916000520

Hollien, H., Mendes-Schwartz, A.-P., & Nielsen, K. (2000). Perceptual confusions of high-pitched sung vowels. *Journal of Voice, 14*, 287-298. https://doi.org/10.1016/S0892-1997(00)80038-7

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology, 39*(3-4), 159-207. https://doi.org/10.1006/cogp.1999.0716

Kitamura, C., & Notley, A. (2009). The shift in infant preferences for vowel duration and pitch contour between 6 and 10 months of age. *Developmental Science, 12*(5), 706-714. https://doi.org/10.1111/j.1467-7687.2009.00818.x.

Ko, E.S., Soderstrom, M., & Morgan, J. (2009). Development of perceptual sensitivity to extrinsic vowel duration in infants learning American English. *Journal of the Acoustical Society of America, 126*(5), 134-139. https://doi.org/10.1121/1.3239465

Koelsch, S., & Siebel, W. A. (2005). Towards a neural basis of music perception. *Trends in Cognitive Sciences, 9,* 578–584. https://doi.org/10.1016/j.tics.2005.10.001

Krämer, M. (2009). *The Phonology of Italian*. Oxford University Press.

Kuhl, P. K., Andruški, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science, 277*(5326), 684–686. https://doi.org/10.1126/science.277.5326.684

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 979-1000. https://doi.org/10.1098/rstb.2007.2154

Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: Implications for the "critical period". *Language Learning and Development*, *1*(3-4), 237-264. https://doi.org/10.1080/15475441.2005.9671948

Kuhl, P. K., Ramírez, R. R., Bosseler, A., Lin, J.-F. L., & Imada, T. (2014). Infants' brain responses to speech suggest analysis by synthesis. *Proceedings of the National*

Academy of Sciences, 111(31), 11238-11245. https://doi.org/10.1073/pnas.1410963111

Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, *100*(15), 9096-9101. https://doi.org/10.1073/pnas.1532872100

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 31*(255), 606-608. doi: 10.1126/science.1736364

Lebedeva, G. C., & Kuhl, P. K. (2010). Sing that tune: Infants' perception of melody and lyrics and the facilitation of phonetic recognition in songs. *Infant Behavior and Development*, *33*(4), 419-430. https://doi.org/10.1016/j.infbeh.2010.04.006

Liu, H. M., Kuhl, P. K., & Tsao, F. M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science*, *6*(3), F1-F10. https://doi.org/10.1111/1467-7687.00275

Liu, H. M., Tsao, F. M., & Kuhl, P. K. (2007). Acoustic analysis of lexical tone in Mandarin infant-directed speech. *Developmental Psychology*, *43*(4), 912-917. https://doi.org/10.1037/0012-1649.43.4.912

London, J. (2012). *Hearing in time: Psychological aspects of musical meter.* Oxford University Press.

Ludusan, B., Cristia, A., Martin, A., Mazuka, R., & Dupoux, E. (2016). Learnability of prosodic boundaries: Is infant-directed speech easier? *The Journal of the Acoustical Society of America*, *140*(2), 1239. https://doi.org/10.1121/1.4960576

Malloch, S. N. (1999). Mothers and infants and communicative musicality. *Musicæ Scientiæ, 3*(1), 29-57. https://doi.org/10.1177/10298649000030S104

Malloch, S., and Trevarthen, C. (2009). Musicality: communicating the vitality and interests of life. In S. Malloch and C. Trevarthen (Eds), *Communicative Musicality: Exploring the Basis of Human Companionship* (pp.1-11). Oxford University Press.

Martin, A., Igarashi, Y., Jincho, N., & Mazuka, R. (2016). Utterances in infant-directed speech are shorter, not slower. *Cognition*, *156*, 52-59. https://doi.org/10.1016/j.cognition.2016.07.015

Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., & Cristia, A. (2015). Mothers speak less clearly to infants than to adults: a comprehensive test of the hyperarticulation hypothesis. *Psychological Science*, *26*(3), 341-347. https://doi.org/10.1177/0956797614562453

Mattock, K., Molnar, M., Polka, L., and Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition, 106*, 1367-1381. https://doi.org/10.1016/j.cognition.2007.07.002

Maurer, D., & Werker, J. F. (2014). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology, 56*(2), 154-178. https://doi.org/10.1002/dev.21177

McCarthy, K. M., Skoruppa, K., & Iverson, P. (2019). Development of neural perceptual vowel spaces during the first year of life. *Scientific Reports*, *9*(1), 19592. https://doi.org/10.1038/s41598-019-55085-y

Mehr, S. A., & Spelke, E. S. (2018). Shared musical knowledge in 11-month-old infants. *Developmental Science*, *21*(2), e12542. https://doi.org/10.1111/desc.12542

Mehr, S. A., Singh, M., York, H., Glowacki, L., and Krasnow, M. M. (2018). Form and function in human song. *Current Biology, 28*(3): 356-368.e5. https://doi.org/10.1016/j.cub.2017.12.042

Missaglia, F. (2004). Distorted perception and production of L2-German vowels by adult Italian learners. In J. Drevillon, J. Vivier, & A Salinas (Eds), *Psycholinguistics, A*

Multidisciplinary Science of 2000: What Implications, What Applications? (pp. 305-314). Université de Caen.

Mugitani, R., Pons, F., Fais, L., Dietrich, C., Werker, J.F., & Amano, S. (2009). Perception of vowel length by Japanese- and English-learning infants. *Developmental Psychology, 45*(1), 236-247. https://doi.org/10.1037/a0014043

Nakata, T., & Mitani, C. (2005). Influences of temporal fluctuation on infant attention. *Music Perception, 22*(3), 401-409. https://doi.org/10.1525/mp.2005.22.3.401

Nakata, T., & Trehub, S. E. (2004). Infants' responsiveness to maternal speech and singing. *Infant Behavior and Development*, *27*(4), 455-464. https://doi.org/10.1016/j.infbeh.2004.03.002

Papageorgiou, K. A., Smith, T. J., Wu, R., Johnson, M. H., Kirkham, N. Z., & Ronald, A. (2014). Individual differences in infant fixation duration relate to attention and behavioral control in childhood. *Psychological Science, 25*(7), 1371-1379. https://doi.org/10.1177/0956797614531295

Papoušek, M., Papoušek, H., & Symmes, D. (1991). The meanings of melodies in motherese in tone and stress languages. *Infant Behavior and Development*, *14*(4), 415-440. https://doi.org/10.1016/0163-6383(91)90031-M

Pegg, J. E., Werker, J. F., & McLeod, P. J. (1992). Preference for infant-directed over adult-directed speech: Evidence from 7-week-old infants. *Infant Behavior and Development, 15*(3), 325-345. https://doi.org/10.1016/0163-6383(92)80003-D

Piazza, E. A., Iordan, M. C., & Lew-Williams, C. (2017). Mothers consistently alter their unique vocal fingerprints when communicating with infants. *Current Biology*, *27*(20), 3162–3167.e3. https://doi.org/10.1016/j.cub.2017.08.074

Pisoni, D.B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception and Psychophysics, 15*(2), 285-290. https://dx.doi.org/10.3758%2Fbf03213946

Politimou, N., Dalla Bella, S., Farrugia, N., & Franco, F. (2019). Born to speak and sing: Musical predictors of language development in pre-schoolers. *Frontiers in Psychology*, *10*, 948. https://doi.org/10.3389/fpsyg.2019.00948

Politimou, N., Stewart, L., Müllensiefen, D., & Franco, F. (2018). Music@ Home: A novel instrument to assess the home musical environment in the early years. *PloS one*, *13*(4), e0193819. https://doi.org/10.1371/journal.pone.0193819

Polka, L., & Bohn, O.-S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics, 39*(4), 467-478. https://doi.org/10.1016/j.wocn.2010.08.007

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 421-435. https://doi.org/10.1037//0096-1523.20.2.421

Provasi, J., Anderson, D. I., & Barbu-Roth, M. (2014). Rhythm perception, production, and synchronization during the perinatal period. *Frontiers in Psychology*, *5*, 1-16. https://doi.org/10.3389/fpsyg.2014.01048

Pruitt, J. S., Jenkins, J. J., & Strange, W. (2006). Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *The Journal of the Acoustical Society of America*, *119*(3), 1684–1696. https://doi.org/10.1121/1.2161427

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2014). Look who's talking: speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science, 17*(6), 880-891. https://doi.org/10.1111/desc.12172

Räsänen, O., Kakouros, S., & Soderstrom, M. (2018). Is infant-directed speech interesting because it is surprising? Linking properties of IDS to statistical learning and attention at the prosodic level. *Cognition, 178*, 193-206. https://doi.org/10.1016/j.cognition.2018.05.015

Sato, Y., Sogabe, Y., & Mazuka, R. (2010). Discrimination of phonemic vowel length by Japanese infants. *Developmental Psychology, 46*(1), 106-119. https://doi.org/10.1037/a0016718

Savy, R. & Cutugno, F. (1997). Ipoarticolazione, riduzione vocalica, centralizzazione: come interagiscono nella variazione diafasica. In F. Cutugno (Ed.), *Fonetica e fonologia degli stili dell'italiano parlato: atti delle 7. Giornate di studio del Gruppo di fonetica sperimentale A.I.A.: Napoli, 14-15 novembre 1996*, *Roma* (pp. 177-194). Esagrafica.

Schellenberg, E. G., & Trehub, S. E. (1996). Natural musical intervals: Evidence from infant listeners. *Psychological Science, 7*(5), 272-277. https://doi.org/10.1111/j.1467-9280.1996.tb00373.x

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. Https://doi.org/10.1016/j.cognition.2007.03.005

Seidl, A., Cristià, A., Bernard, K.H., & Onishi, K.H. (2009). Allophonic and phonemic contrasts in infants' learning of sound patterns. *Language Learning and Development, 5*(19), 191-202. https://doi.org/10.1080/15475440902754326

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology*, *18*(9), 668-667. https://doi.org/10.1016/j.cub.2008.03.059

Singh, L., Morgan, J. L., & Best, C. T. (2002). Infants' listening preferences: Baby talk or happy talk? *Infancy, 3*, 365–394. https://doi.org/10.1207/S15327078IN0303_5

Snijders, T. M., Benders, T., & Fikkert, P. (2020). Infants segment words from songs—An EEG study. *Brain Sciences*, *10*(1), 39. https://doi.org/10.3390/brainsci10010039

Soderstrom, M. (2019). ManyBabies1: Infants' preference for infant-directed speech. *The Journal of the Acoustical Society of America, 145(*3), 1728-1728. https://doi.org/10.1121/1.5101348

Soley, G., & Hannon, E. E. (2010). Infants prefer the musical meter of their own culture: a cross-cultural comparison. *Developmental Psychology*, *46*(1), 286. https://doi.org/10.1037/a0017555

Spinelli, M., Fasolo, M., & Mesman, J. (2017). Does prosody make the difference? A meta-analysis on relations between prosodic aspects of infant-directed speech and infant outcomes. *Developmental Review*, *44*, 1– 18. https://doi.org/10.1016/j.dr.2016.12.001

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*(6640), 381-382. https://doi.org/10.1038/41102

Sundberg, J. (1987). *The Science of the Singing Voice*. Northern Illinois University Press.

Sundberg, J. (1989). Synthesis of singing by rule. In M. V. Mathews, & J. R. Pierce (Eds.), *Current directions in computer music research* (pp.45-55). MIT Press.

Sundberg, J. (2001). Level and center frequency of the singer's formant. *Journal of Voice*, *15*, 176-186. https://doi.org/10.1016/S0892-1997(01)00019-4

Sundberg, J., & Romedahl, C. (2009). Text intelligibility and the singer's formant – a relationship? *Journal of Voice, 23*, 539-545. https://doi.org/10.1016/j.jvoice.2008.01.010

Ter Schure, S., Junge, C., & Boersma, P. (2016). Discriminating non-native vowels on the basis of multimodal, auditory or visual information: Effects on infants' looking patterns and discrimination. *Frontiers in Psychology, 7*, 1-11. https://doi.org/10.3389/fpsyg.2016.00525

Thiessen, E. D., & Saffran, J. R. (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. https://doi.org/10.1111/j.1749-6632.2009.04547.x

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*(1), 53-71. https://doi.org/10.1207/s15327078in0701_5

Trainor, L. J. (1996). Infant preferences for infant-directed versus non infant-directed playsongs and lullabies. *Infant Behavior and Development*, *19*(1), 83-92. https://doi.org/10.1016/S0163-6383(96)90046-6

Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin & Review*, *9*(2), 335-340. https://doi.org/10.3758/BF03196290

Trainor, L. J., & Trehub, S. E. (1992). A comparison of infants' and adults' sensitivity to Western musical structure. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(2), 394-402. https://doi.org/10.1037/0096-1523.18.2.394

Trainor, L. J., Clark, E. D., Huntley, A., & Adams, B. A. (1997). The acoustic basis of preferences for infant-directed singing. *Infant Behavior and Development*, *20*(3), 383-396. https://doi.org/10.1016/S0163-6383(97)90009-6

Trehub, S. E., & Gudmundsdottir, H. R. (2015). Mothers as singing mentors for infants. In G. F. Welch, D. M. Howard, & J. Nix (Eds.), *Oxford Handbook of Singing (pp. 1-12).* Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199660773.013.25

Trehub, S. E., & Hannon, E. E. (2009). Conventional rhythms enhance infants' and adults' perception of musical patterns. *Cortex, 45*(1), 110-118. https://doi.org/10.1016/j.cortex.2008.05.012

Trehub, S. E., & Trainor, L. (1998). Singing to infants: Lullabies and play songs. *Advances in Infancy Research*, *12*, 43-78.

Trehub, S. E., Bull, D., & Thorpe, L. A. (1984). Infants' perception of melodies: The role of melodic contour. *Child Development, 55*(3), 821-830. https://doi.org/10.2307/1130133

Trehub, S. E., Ghazban, N., & Corbeil, M. (2015). Musical affect regulation in infancy. *Annals of the New York Academy of Sciences*, *1337*(1), 186-192. https://doi.org/10.1111/nyas.12622

Trehub, S. E., Plantinga, J., & Russo, F. A. (2016). Maternal vocal interactions with infants: Reciprocal visual influences. *Social Development*, *25*(3), 665-683. https://doi.org/10.1111/sode.12164

Trehub, S.E., Becker, J., & Morley, I. (2015). Cross-cultural perspectives on music and musicality. *Philosophical Transactions of the Royal Society, B 370*: 20140096. https://doi.org/10.1098/rstb.2014.0096

Trehub, S.E., Unyk, A.M., & Trainor, L.J. (1993). Maternal singing in cross-cultural perspective. *Infant Behavior and Development, 16*, 285-295. https://doi.org/10.1016/0163-6383(93)80036-8

Tsang, C. D., Falk, S., & Hessel, A. (2017). Infants prefer infant-directed song over speech. *Child Development*, *88*(4), 1207-1215. https://doi.org/10.1111/cdev.12647

Tsao, F. M., Liu, H. M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development*, *75*(4), 1067-1084. https://doi.org/10.1111/j.1467-8624.2004.00726.x

Tsuji, S., & Cristia, A. (2014). Perceptual attunement in vowels: A meta-analysis. *Developmental Psychobiology, 56*(2), 179-191. https://doi.org/10.1002/dev.21179

Van der Heijden, M. J., Araghi, S. O., Jeekel, J., Reiss, I. K., Hunink, M. M., & Van Dijk, M. (2016). Do hospitalized premature infants benefit from music interventions? A systematic review of randomized controlled trials. *PloS One*, *11*(9), e0161848. https://doi.org/10.1371/journal.pone.0161848

Van Puyvelde, M., & Franco, F. (2015). The interaction of music and language in the ontogenesis of human communication: a multimodal parent-infant co-regulation system. In Timmers, R., Dibben, N., Eitan, Z., Granot, R., Metcalfe, T., Schiavio, A. & Williamsom, V. (Eds). *Proceedings of ICMEM International Conference on the Multimodal Experience of Music 2015* (pp.1-8). HRI Online Publications. ISBN 9780957102248.

Volkova, A., Trehub, S. E., & Schellenberg, E. G. (2006). Infants' memory for musical performances. *Developmental Science, 9*(6), 583-589. https://doi.org/10.1111/j.1467-7687.2006.00536.x

Weirich, M., & Simpson, A. (2019). Effects of gender, parental role, and time on infant- and adult-directed read and spontaneous speech. *Journal of Speech, Language, and Hearing Research*, *62*(11), 4001-4014. https://doi.org/10.1044/2019_JSLHR-S-19-0047

Weisleder, A., & Fernald, A. (2013). Talking to children matters: early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143-2152. https://doi.org/10.1177/0956797613488145

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development, 1*(2), 197-234. https://doi.org/10.1207/s15473341lld0102_4

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49-63. https://doi.org/10.1016/S0163-6383(84)80022-3

Werker, J. F., & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology, 46*(3), 233-251. https://doi.org/10.1002/dev.20060

Werker, J. F., Gilbert, J. H. V., Humphrey, G. K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, *52*, 349–355. https://doi.org/10.2307/1129249

Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., & Honing, H. (2009). Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences*, *106*(7), 2468-2471. https://doi.org/10.1073/pnas.0809035106

Yu, C., Suanda, S. H., & Smith, L. B. (2019). Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Developmental Science, 22*(1), e12735. https://doi.org/10.1111/desc.12735

Zhang, Y., Kuhl, P. K., Imada, T., Kotani, M., & Tohkura, Y. I. (2005). Effects of language experience: neural commitment to language-specific auditory patterns. *NeuroImage*, *26*(3), 703-720. https://doi.org/10.1016/j.neuroimage.2005.02.040

Zhao, T. C., & Kuhl, P. K. (2016). Musical intervention enhances infants' neural processing of temporal structure in music and speech. *Proceedings of the National Academy of Sciences*, *113*(19), 5212-5217. https://doi.org/10.1073/pnas.1603984113

Tables & Figures

Table 1. Description of the participants' sample in Experiment 1.

	Speech	Singing
N	12	12
Infant Gender	8 F, 4 M	7 F, 5 M
Mean age (SD)	4.8 (0.72)	4.6 (0.5)
Exposure to singing (SD)*	30.58 (4.07)	31.25 (3.54)
Maternal education (SD)**	15.17 (2.85)	13.83 (2.88)

*Exposure to singing score is derived from a subscale (questions 14-18) in the Music@Home questionnaire (Politimou et al., 2018).

**Maternal education is measured in years of school education from primary school.

Table 2. Acoustic features of speech and song stimuli. Mean and SD (in brackets).

	Speech	Singing
Pitch range (Hz)	176-517	285-414
Intersyllable interval (ms)	406 (27)	409 (20)
Vowel duration (ms)	220 (56)	224 (109)
F1 /u/ (Hz)	390 (60)	412 (56)
F2 /u/ (Hz)	909 (96)	985 (130)
F1 /y/ (Hz)	404 (113)	403 (90)
F2 /y/ (Hz)	2103 (67)	2166 (31)

Figure 1

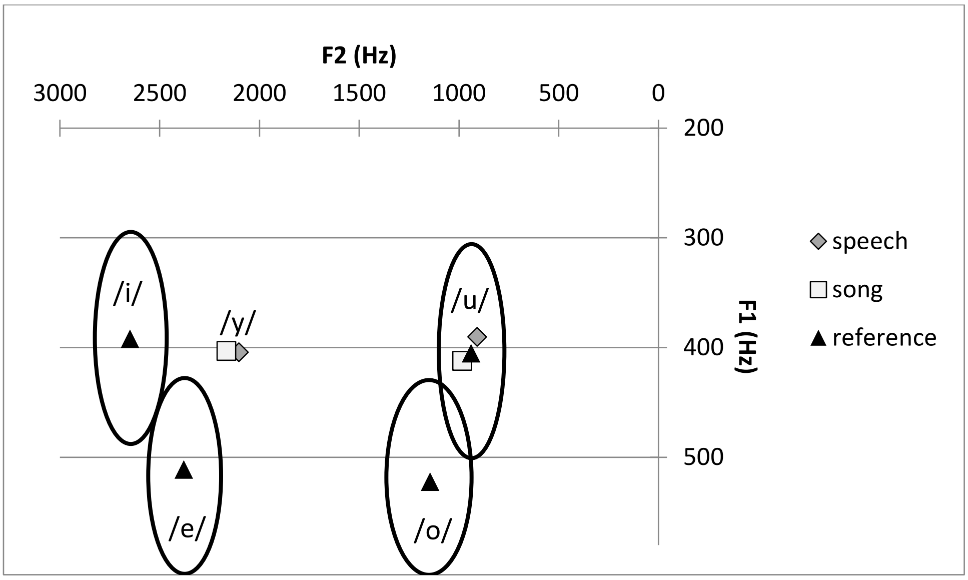

Figure 1: Mean first (F1, y-axis) and second (F2, x-axis) formant values of the vowels /y/ and /u/ used in the experiment (light grey square: song; dark grey diamond: speech), in relation to reference values (black triangles, adopted from Fig.4 in Savy & Cutugno, 1997) for /u/ and other nearby vowel categories in conversational Italian of female speakers in Northern Italy (Lombardia) showing the median and 1 SD ellipsis for each vowel category.

Figure 2

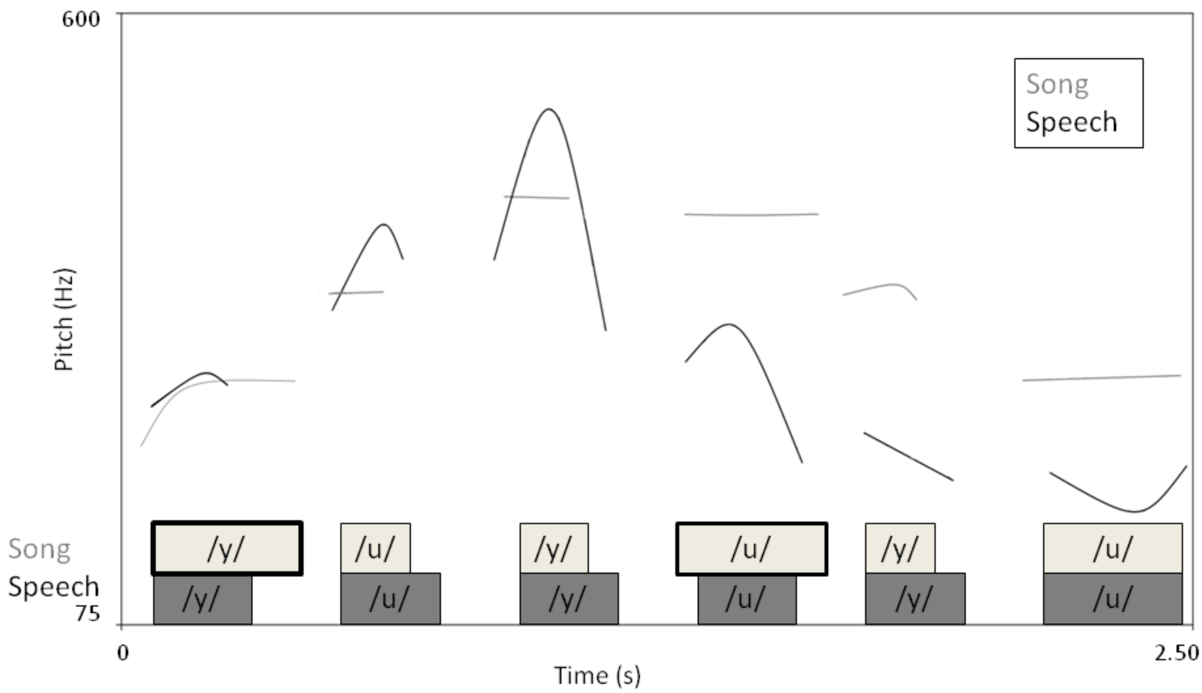

Figure 2: Pitch (Hz, lines) and duration (s, squares) of vowels in varied ID-like speech (black) and song (grey) syllable sequence. Pitch contours were smoothed to give a clearer view of the overall contour. Bold squares indicate metrically strong vowels in the song condition. Note that the duration of the final syllables in both ID-like speech and song is lengthened due to its phrase-final position.

Figure 3

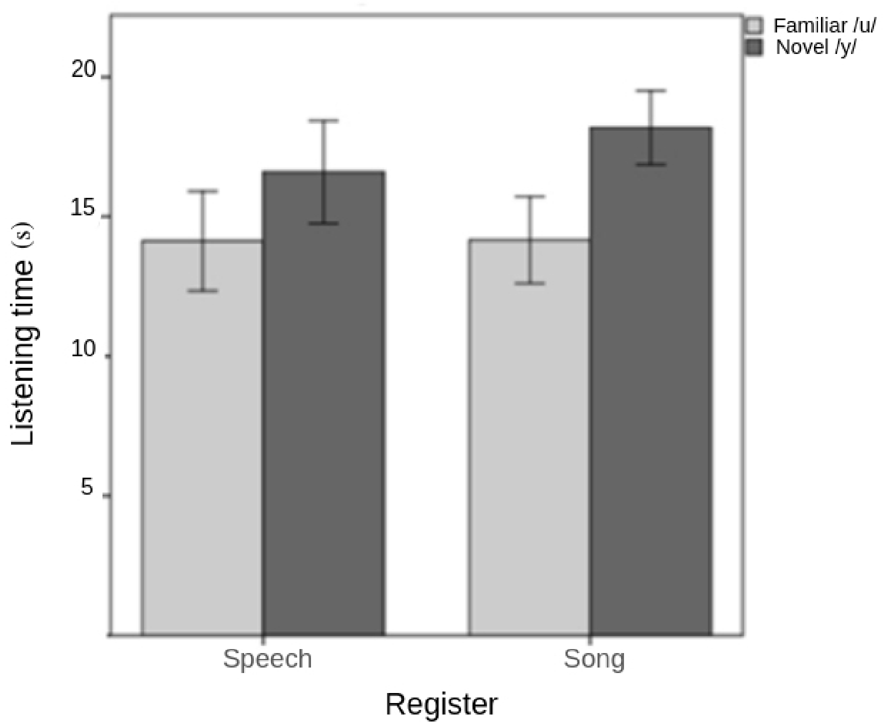

Figure 3. Four-to-six-month-old infants' mean listening times (sec) during non-alternating sequences (familiar vowel only) and alternating sequences (including novel vowel) presented in speech and song contexts. Error bars show the Standard Error of the mean.

Table 3. Description of the participants' sample in Experiment 2.

	Speech	Singing
N	12	12
Gender	6 F, 6 M	6 F, 6 M
Mean age (SD)	8 (0.27)	8.7 (0.72)
Exposure to singing (SD)*	28.25 (4.88)	29 (4.32)
Maternal education (SD)**	14.83 (2.69)	15.17 (2.69)

*Exposure to singing score is derived from a subscale (questions 14-18) in the Music@Home questionnaire (Politimou et al., 2018).

**Maternal education is measured in years of school education from primary school.

Figure 4

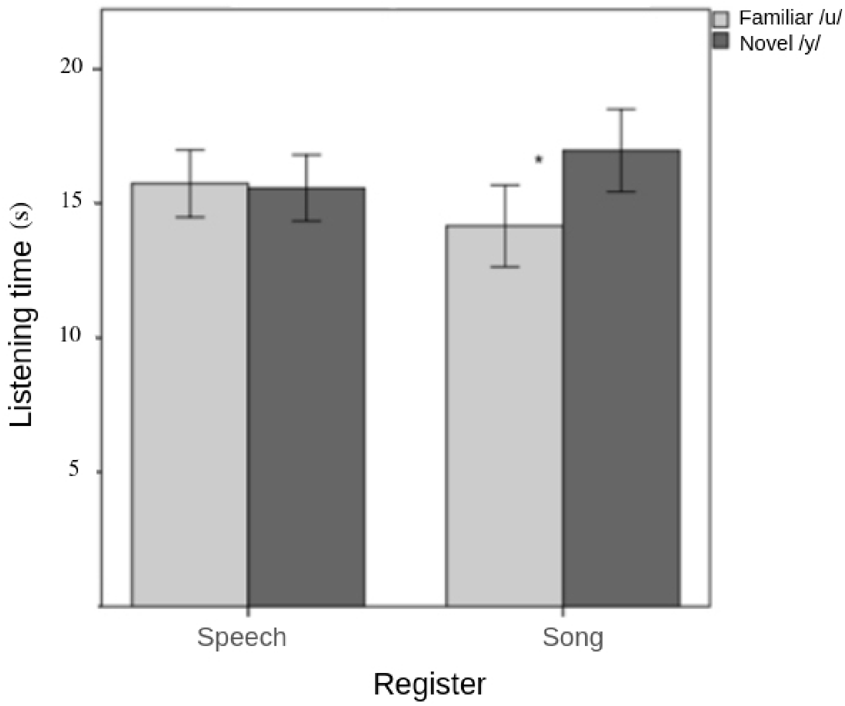

Figure 4. Eight to ten months old infants' mean listening times (sec) during non-alternating (familiar vowel) and alternating sequences (including novel vowel) presented in speech and song contexts. Error bars show the Standard Error of the mean.