Adults' and infants' perception of infant-directed speech and song

Simone Falk[1,2], Christine D. Tsang[3]
[1]Laboratoire Parole et Langage, Aix-Marseille Université, CNRS, UMR 7309, Aix-en-Provence, France
[2]Department of German Philology, Ludwig-Maximilians-University, Munich, Germany
[3]Department of Psychology, Huron College at Western, London, Ontario, Canada
simone.falk@germanistik.uni-muenchen.de, ctsang33@huron.uwo.ca

ABSTRACT

Infant-directed speech, also sometimes described as "musilanguage", has acoustic characteristics overlapping with those of infant-directed singing. In the present contribution we determine whether native-English-speaking adults and infants from native-English speaking households discriminate between infant-directed speech and song in a non-native language and which acoustic features are typical of their speech or song perception. Twenty-four native English-speaking adults rated Russian speech and song samples taken from naturalistic infant-directed recordings. Furthermore, 32 infants (6 to 9 months) were tested in a head-turn preference procedure on a selection of the stimuli rated by adults. Results showed that both infant and adults discriminated between infant-directed speech and singing in a foreign language. In adults, tempo and pitch variability accounted for perceived differences. Moreover, infants, independently of age, attended longer to song than to speech – an intriguing result underlining the importance of studying infant-directed singing and its role for early linguistic development.

Keywords: Infant-directed communication; speech and song; adult and infant perception.

1. INTRODUCTION

Humans readily distinguish between speech and musical signals. However, the boundaries between speech and song are less clear, sometimes speech can even be perceived as song [4]. In infant-directed (ID) communication, caregivers of diverse cultures and linguistic backgrounds have been found to modify pitch and rhythmic characteristics of their spoken utterances in a musical way when addressing infants and young children [9, 17]. Another frequent, but less studied, part of ID-communication is ID-singing. Singing and speaking with infants share multiple acoustic characteristics when compared to adult-directed singing and speaking [8, 16, 19]. Higher pitch, slower tempo, shorter utterances, prominent intervals, clearer speech quality and a more smily voice are among the typical attributes of both ID-registers [7, 18]. However, less is known when it comes to the acoustic and perceptual differences between ID-speech and ID-singing. In the present study, we examine the question of whether ID-speech and ID–singing are discriminated by adult and infant listeners and, if so, which acoustic characteristics are involved in this discrimination. Therefore, we chose samples taken from naturalistic ID-singing and ID-speaking contexts in an unfamiliar language and cultural background and presented them to Canadian English-speakers, adults and infants in their first year of life. For adults, we expect clear discrimination of both registers along the lines of acoustic differences between non-ID speech and song described in the literature [6]. However, infants may not yet be able to differentiate ID-speech and ID-song stimuli [2] as the prosodic characteristics of both registers serve to communicate global pragmatic and emotional messages at young age [14, 15] which may be highly relevant for infants beyond the linguistic or musical specificity of the sound.

2. ADULT PERCEPTION

In the first part of the study, we aimed at determining acoustic features which relate to adults' perception of unfamiliar ID-speech and ID-song.

2.1. Participants

Twenty-four English-speaking adults (mean age = 18 years; age range 17 years - 19 years; 3 male), enrolled in a first year undergraduate Psychology course at a university in London Ontario Canada participated in the study. All participants indicated that English was their native language.

2.2. Material

Forty-nine short excerpts were chosen from a corpus of naturalistic Russian infant-directed speech and singing [5]. 10 Russian mothers (mean age = 32.0 years, *SD* = 4.5) were speaking and singing in the presence of their infants (mean age = 7.9 months, *SD* = 3.7). The excerpts were comparable in mean number of syllables (singing: 7.4, *SD* = 2.0; speech: 7.4, *SD* = 3.9), mean pitch (singing: 284 Hz, *SD* = 62.3 Hz; speech: 292, *SD* = 66.4 Hz) and were similar with respect to the shapes of their global

melodic contour. 24 of the excerpts (mean length = 2.65 s, *SD* = .62) were taken from musical contexts, that is, when mothers sang traditional Russian playsongs and lullabies to their infants. Twenty-five excerpts were taken from typical speech contexts, that is, from mother-infant dialogues and from traditional Russian rhymes or children's poems (without melodic tradition) that mothers recited during play (mean length = 1.99 s, *SD* = .74).

2.3. Acoustic analyses

We measured the following parameters for each excerpt in order to correlate them with the obtained adult ratings: the duration of vowels and consonants and the derived durational portion of vowels in percent, mean tempo (syllables per second), the range from the f0 minimum to f0 maximum of each excerpt ("overall pitch range") expressed in semitones and the range from f0 minimum to f0 maximum in each vowel ("pitch variability in vowels") in semitones, the duration of intervocalic intervals and the coefficient of variation of these intervals. All pitch and duration analyses were carried out using Praat [1].

2.4. Procedure

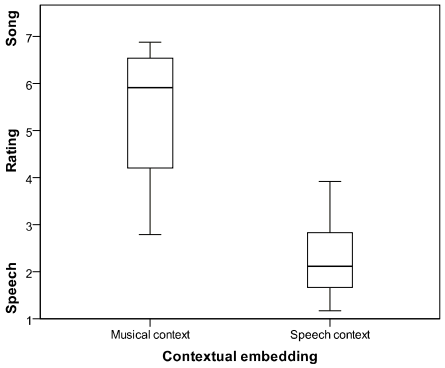
Participants listened twice to each of the 49 short ID excerpts presented via loudspeakers connected to a laptop computer in a quiet room. Participants were asked to rate each sample on a 7-point Likert-scale (1= "Definitely Speech" and 7 = "Definitely Singing"). There were two random orders of presentation, and participants were randomly assigned to one order condition.

2.5. Results and Discussion

English-speaking adults' ratings differed significantly for Russian ID-samples taken from musical and speech contexts ($t(35.79) = 9.5$, $p < .001$, see Figure 1). Only 10 out of the overall 49 excerpts were rated in the ambiguous (neither speech nor song) range between 3-5.

We determined which acoustic properties were most related to the English-speaking adults' perception of Russian ID-speech and ID-song. Acoustic features (tempo, pitch range, mean pitch stability, mean duration and percent of vowels, mean duration of consonants and variability of intervocalic intervals) were correlated to the adult speech/song ratings on the excerpts. Results are displayed in Table 1 (higher ratings are typical of song perception). Consonant duration and variability of intervocalic intervals were the only properties that were not correlated to the ratings.

Figure 1: Adult ratings of infant-directed speech and song samples).



Samples with slower tempo, smaller pitch range, smaller pitch variability as well as more and longer vowel portions tended to be rated more often as song.

Table 1: Correlations between adult ratings and acoustic features of ID-speech and -song samples.

Measure	Ratings
Tempo (syll/s)	-.50***
Range (st)	-.47**
Pitch variability (vowels, st)	-.45**
Duration (vowels, s)	.47**
Duration (consonants, s)	.10
Vowel portion (%)	.44**
Intervocalic intervals variability	.05

*$p < .05$ **$p < .01$, ***$p < .001$

At step 1 of the analysis, Tempo (syll / s) was significantly related to the ratings (Adjusted *R* square = .23, $F(1,47) = 15.29$, $p < .001$). At step 2, Tempo and Pitch Variability entered the equation and were significantly related to the rating (Adjusted *R* square = .483, $F(2,46) = 23.40$, $p < .001$, Tempo standardized $\beta = -.56$, $t = -3.91$, $p < .001$; Pitch Variability standardized $\beta = -.51$, $t = -4.9$, $p < .001$). The other predictors did not enter the equation in further steps. Hence, approximately 48% of the variability of the rating data could be explained by two acoustic variables.

In sum, adults easily distinguished between ID-speech and -song contexts, as expected. In particular, slower tempo and vowel properties, in particular, decreased pitch variability (i.e., more stable tonal targets), were interpreted as cues to ID-"songness" compared to ID-"speechness".

3. INFANT PERCEPTION

In the second part of the study, we examined whether infants' perception differed for ID-speech and ID-song using a preference procedure. In two

previous studies [2, 3] with infants in the second half year of life, no preference could be established for either speech or song in a non-native language, when stimuli were presented only auditorily. However, the stimuli used in these studies consisted of one song sample and one speech sample that were repeated several times. Moreover, the singer/speaker mimicked ID-speech and ID-song styles (i.e, the stimuli were not truly infant-directed), and the speech sample was always derived from the song sample. We expect that, when using naturalistic ID-recordings whose acoustics were most characteristic of speech and song for adult listeners, infants would be more likely to differentiate between both registers provided that discriminating between these two types of stimuli is relevant to them.

3.1. Participants

Sixteen 6- to 7-month-old infants (eight males, mean age = 7.1 months, range = 6.0-7.9 months) and sixteen 8- to 9-month old infants (eight males, mean age = 9.4, range = 8.8-9.9 months) participated in the experiment.

3.2. Material

Based on the adult study, samples were chosen to constitute the ID-speech and ID-song stimuli that were most unanimously rated as speech or song. To create a "Speech Set", six ID-speech samples from four different mothers were concatenated that had a mean rating of 1.5 or less (σ = 0.15 – 1.01) in the adult study. Similarly, to create a "Song Set", the five ID-song samples with the highest rating in the adult study (mean rating of 6.5 or higher, σ = 0.15 – 0.87) were concatenated. Acoustic features that differed significantly (Mann-Whitney-Tests, $p < .05$) in both sets are displayed in Table 2.

Table 2: Acoustic differences between ID-song and ID-speech stimuli chosen from adults' top ratings (Standard deviations in brackets).

Measure	ID-song	ID-speech
Tempo (syll/s)	2.5 (0.60)	4.6 (1.16)
Range (st)	8.0 (4.0)	19.8 (6.18)
Pitch variability (st)	1.6 (0.33)	2.9 (0.78)
Vowel duration (ms)	340 (45)	222 (26)
Vowel portion (%)	63.4 (6.8)	44.5 (5.8)

Each set lasted approximately 12 seconds. Five versions of each set were created by randomizing the order of the samples in the sequence. These versions were concatenated with an intern-onset interval of 50 ms in quasi-random order, such that the same speech/song sample did not repeat in succession. This procedure resulted in two final stimulus sets of ID-speech and ID-song lasting 1 min each.

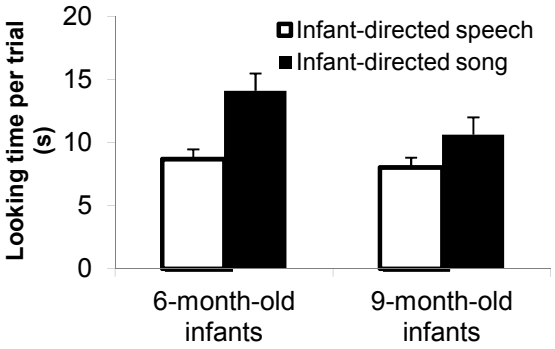
3.3. Procedure

Infants were tested individually in a head-turn preference procedure, seated on their parents' lap in a quiet laboratory room [10]. The speech and song stimuli were played by loudspeakers situated on either the left or the right side of the infant. As soon as the infant turned his/her head to face a visual stimulus on a monitor below the speaker, the auditory stimulus started (e.g., the ID-song stimulus set). During the trial, the 1 min-stimulus set played continuously on a loop until the infant looked away (45 degree head turn) for at least 2 sec. In the following trial, the other stimulus set (e.g., ID-speech) was presented on the opposite side of the infant. Thus, speech and song trials alternated throughout the experiment until the infant had completed 20 trials in total. Infants were counterbalanced for the side and stimulus set with which they started.

3.4. Results & Discussion

Results are displayed in Figure 2. Using looking time (in seconds) as the dependent variable, we conducted a 2 x 2 analysis of variance with Stimulus (ID-speech or ID-song) as the within-subjects variable and Age (6-months or 9-months) as the between-subjects variable. Independently of age, infants' looking times were significantly longer for ID-song than for ID-speech as indicated by a main effect of Stimulus ($F(1,30) = 12.69$, $p = .001$).

Figure 2: Infants' looking times for ID-speech and song (error bars display standard errors).



No interaction was found for Age and Stimulus ($F(1,30) = 1.55$, $p = 0.22$). A marginal main effect was found for Age ($F(1,30) = 3.51$, $p = .07$), such that 6-month-old infants looked slightly longer overall than 9-month-old infants.

In contrast to previous studies [2, 3], our findings showed that 6- to 9-month-old English-learning infants were more attracted to ID-song compared to ID-speech in a non-native language. These ID-stimuli were judged by English-speaking adults as being most typical instances of speech and song. The observation of a preference for ID-song is consistent with results from Nakata & Trehub [13] who tested infants' preference for their mothers singing or speaking to them in their native language in an infant-directed and multimodal setting (i.e., infants saw a visual display of their mother's face and heard her voice simultaneously). The finding also confirms that infants in their first year of life are able to discriminate between both registers.

As there were multiple acoustic differences between both stimulus sets, it is left to future research to establish which of the acoustic characteristics (see Table 2) are critical to infants' perception of ID-stimuli, and more importantly, what acoustic characteristics allow for the perceptual discrimination between ID-speech and ID-song. Still, we may speculate about some aspects known to influence infants' perception. Acoustic analyses in the first part of the study revealed that ID-singing most strikingly affects the segmental organisation of speech. It makes vocalic portions of the signal physically more prominent and distinct. In fact, 6-month-old infants are attracted towards prolonged vowels in utterances, while 10-month-olds already prefer to listen to vowels that have durations more typical of adult speech [11]. Hence, it is a possibility that infants preferred to listen to ID-song compared to ID-speech because of its more prominent vocalic portions.

Moreover, the combination of acoustic qualities in ID-singing (e.g., longer vowels with higher pitch stability, more constrained pitch range and slower tempo) may be helpful to facilitate the extraction of global tonal contours by young listeners. Contour information is highly indicative of communicative and emotional messages in ID-communication [7, 15]. Infants at a young age are very sensitive to this information [8]. Our age group, between 6 and 9 months, may still rely on and be more attracted to this information than older infants who are more immersed in linguistic interpretation of their native language prosody (e.g., as a means of word learning). In sum, the acoustic characteristics of ID-singing may be highly adapted to infants' developmental progress at 6 to 9 months which may account for the attentional preference at this age.

Finally, in our study, infants were exposed to an unfamiliar language. It is known that infants in their second half year perceptually tune in to the segmental and prosodic structure of their native language [12, 20]. Thus, they may lose interest in speech signals that do not conform to their "expectations" with respect to segmental variation. Still, it is an open question why ID-singing would not undergo the same process of perceptual narrowing. In sum, future studies should examine preferences at older and younger ages as well as infant behaviour when confronted with ID-speech and -singing in their native language.

3. CONCLUSION

Adults and infants listen differently to ID-speech and ID-song stimuli. For infants, we showed for the first time that, relying on the auditory signal alone, 6 to 9 month olds prefer to listen to ID-singing. This result underlines the relevance of ID-singing as a basic form of infant-communication which should be studied in its effects on infant development and language acquisition. As for adults, perceived differences between ID-speech and ID–song may derive from categorical judgments on the acoustics with sung acoustics referring to a musical domain and speech to a language domain. A developmental perspective on ID-song and ID-speech perception may allow for a better understanding of the developmental pathway leading to these highly developed cognitive functions that are essential in human culture and life.

Acknowledgements

This research was supported by the People Program (Marie Curie Actions) of the European Union's Seventh Framework Program (FP7/2007-2013) under REA grant agreement n°327586 to SF, a SSHRC-MCRI AIRS research grant to CDT and SF and a FASS Research Grant from Huron University College to CDT.

7. REFERENCES

[1] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot International 5*, 341–345.

[2] Corbeil, M., Trehub, S. E., Peretz, I. 2013. Speech vs. singing: Infants choose happier sounds. *Front. Psychol. 4*, 372.

[3] Costa-Giomi, E., Ilari, B. 2014. Infants' preferential attention to sung and spoken stimuli. *J. Res. Music Edu. 622*, 188–194.

[4] Deutsch, D., Henthorn, T., Lapidis, R. 2011. Illusory transformation from speech to song. *J. Acoust. Soc. Am. 129*, 2245–2252.

[5] Falk, S. 2009. *Musik und Sprachprosodie. Kindgerichtetes Singen im frühen Spracherwerb*. New York, Berlin : de Gruyter.

[6] Falk, S., Rathcke, T., Dalla Bella, S. 2014. When speech sounds like music. *J. Exp. Psychol. Hum. Percept. Perform. 40*, 1491–1506.
[7] Fernald, A. 1989. Intonation and communicative content in mothers' speech to infants: Is the melody the message? *Child Dev. 60,* 1497–1510.
[8] Fernald, A., Kuhl, P. K. 1987. Acoustic determinants of infant preference for motherese speech. *Infant Behav. Dev, 10*, 279–293.
[9] Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., Fukui, I. 1989. A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *J. Child Lang. 163*, 477–501.
[10] Kemler-Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., Gerken, L. 1995. The head-turn preference procedure for testing auditory perception. *Infant Behav. Dev. 18*, 111–116.
[11] Kitamura, C., Notely, A. 2009. The shift in infant preferences for vowel duration and pitch contour between 6 and 10 months of age. *Dev. Sci. 12*, 706–714.
[12] Maurer, D., Werker, J.F. 2014. Perceptual narrowing during infancy: A comparison of language and faces. *Dev. Psychobiol. 56*, 154–178.
[13] Nakata, T., Trehub, S. E. 2004. Infants' responsiveness to maternal speech and singing. *Infant Behav. Dev. 27*, 455–464.
[14] Nazzi, T., Floccia, C., Bertoncini, J. 1998. Discrimination of pitch contours by neonates. *Infant Behav. Dev. 214*, 779–784.
[15] Papoušek, M., Papoušek, H., Symmes, D. 1991. The meanings of melodies in motherese in tone and stress languages. *Infant Behav. Dev. 14*, 415–440.
[16] Trainor, L. J. 1996. Infant preferences for infant-directed versus non infant-directed playsongs and lullabies. *Infant Behav. Dev. 19*, 83–92.
[17] Trehub, S. E., Trainor, L. 1998. Singing to infants: Lullabies and play songs. *Advances in infancy research 12*, 43–78.
[18] Trehub, S. E., Unyk, A. M., Trainor, L. J. 1993. Adults identify infant-directed music across cultures. *Infant Behav. Dev. 162*, 193–211.
[19] Werker, J. F., Pegg, J. E., McLeod, P. J. 1994. A cross-language investigation of infant preference for infant-directed communication. *Infant Behav. Dev. 173*, 323–333.
[20] Werker, J.F., Tees, R.C. 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev. 71*, 49–63.