PRENATAL INFANT-DIRECTED SPEECH: VOWELS AND VOICE QUALITY

Kateřina Chládková[a,b], Martina Černá[a], Nikola Paillereau[a,b,c], Radek Skarnitzl[a], Zuzana Oceláková[a,1]

[a]Institute of Phonetics, Faculty of Arts, Charles University, [b]Institute of Psychology, Czech Academy of Sciences, [c]Laboratoire de Phonétique et Phonologie (UMR7018, CNRS – Sorbonne Nouvelle)

katerina.chladkova@ff.cuni.cz

ABSTRACT

Speech addressed to young children is typically characterised by distinct acoustic properties. Here we describe infant-directed speech (IDS) addressed to unborn infants. Recordings of pregnant women (speaking Czech) show that prenatal IDS differs from adult-directed speech (ADS) in several vocalic properties. Compared to ADS, prenatal IDS has longer vowels (due to slower speaking rate) with short-long contrasts being preserved. The F1–F2 vowel space is larger in prenatal IDS than in ADS; the vowel-space expansion towards the periphery evokes enhancement of phonemic contrasts. Although overall pitch is not found to differ between styles, in IDS, F0 serves as a supporting cue to short-long contrasts. Voice quality is breathier in IDS than in ADS indicating intimacy or affection. Our results align well with previous findings for 'postnatal' IDS across languages and demonstrate that a distinct infant-directed speech style exists already before birth.

Keywords: prenatal infant directed speech, early speech development, input, Czech, vowels

1. INTRODUCTION

In most language communities adults address young children in a distinct speech style. Infant-directed speech (IDS) differs from adult-directed speech (ADS) in several aspects; for instance, IDS has grammatically simpler and shorter sentences, higher pitch, exaggerated pitch range, slower speech rate and longer pauses [7, 8, 11, 14, 15, 21]. Other, somewhat less frequently studied and reported characteristics of IDS relate to specific phonation qualities, particularly higher breathiness [28].

As for the segmental cues associated with IDS, the literature often focuses on vowel spectral properties. Most studies find the average locations of peripheral vowels such as /a/, /i/ or /u/ to be shifted along the F1 and/or F2 dimension [11]. Some find enlarged vowel spaces in IDS as compared to ADS [6, 22], while other studies do not show any vowel-space exaggeration, and only report higher vowel formants in general [3, 19], larger variation in the realizations of IDS vowels [28], or even a shrinkage of the vowel space in IDS as compared to ADS [3, 12].

The varying outcomes across studies lead to varying proposals as to the function of IDS. Exaggerated F1 and/or F2 distances between vowels have been taken to reflect contrast enhancement facilitating speech-sound acquisition in the language-learning child (cf. [26]). However, shrunk vowel spaces or larger variation in IDS could hardly facilitate speech sound acquisition; some have thus argued that the primary function of IDS might be to communicate positive affect and arouse attention [3, 28].

Consonantal properties in IDS, too, sometimes indicate enlargement of contrast [3, 8] and sometimes could simply reflect different realizations of speech sounds (e.g., longer VOT of voiceless stops which could be a by-product of a slower speaking rate [16, 27]). In languages that contrast phonologically short and long speech sounds, IDS mostly causes lengthening of all segments [11], although in some cases, exaggeration of duration-based contrasts has been reported ([32] for Swedish vowel length, [4] for final consonant voicing in American English).

In sum, IDS exhibits different acoustic properties than ADS, but these may vary across studies and/or languages and may be attributable to various articulatory origins or communicative goals, such as to attract the child's attention, express emotion and facilitate language learning [9].

It is acknowledged that language learning and language-specific perception begin to develop already before birth [10, 17, 29]. Assuming that the function of IDS is, at least partially, to help the developing infant acquire her native language, it would be beneficial if IDS occurred as soon as the (unborn) infant starts to process the speech signal (i.e. from about 25th–28th week of gestational age [18]). To that end, Zhao et al. [33] explicitly suggested that prenatal exposure to IDS, addressed to an older sibling, may enhance perceptual learning already in the womb. The question we ask here is whether IDS is present before birth at all, and if yes, what potential function its acoustic properties might have.

We examine the speech addressed to infants before they are born (prenatal IDS), a phenomenon that has not yet been investigated in developmental literature. If distinctive prenatal IDS exists, it will likely carry acoustic markers of positive affect, such as exaggerated F0, or markers of intimacy, such as softer voice (or breathiness), both of which have previously been reported for IDS [21, 28]. Prenatal IDS could further resemble postnatal IDS in that its acoustic characteristics would support the linguistic development of the unborn child. If any contrast enhancement occurs in prenatal IDS, we expect it to be present for vowel sounds because of their large acoustic salience and supposed developmental precedence over consonants [23]; therefore we focus on vowels.

We assess prenatal IDS in pregnant women speaking Czech as their native language. Considering the widely reported features of 'postnatal' IDS, we test the following. IDS typically displays higher F0 than ADS, which would naturally predict higher F0 in prenatal IDS as well. Note however that F0 in IDS has been shown to vary with infants' age, being lower in speech to newborns than to older infants [21]; the predicted effects on F0 in IDS before birth might thus be rather small or perhaps even not detectable. Enhancement in vowel spectral contrasts, namely the first and the second formant, has been repeatedly – though not always – reported across languages, thus we test whether the vowel space is larger in prenatal IDS than in ADS. Czech is a quantity language contrasting phonologically short and long vowels. Although enhancement of vowel length contrasts is not typically found in IDS, the importance of durational cues (relative to spectral cues) might differ pre- and post-natally; we thus also test whether prenatal IDS differs from ADS in vowel duration.

2. METHOD

2.1. Participants

Recordings were made of 17 female native speakers of Czech (age range 25-41), who were 28+ weeks pregnant (mean = week 33). They were raised in a monolingual Czech environment, and spoke the western (Bohemian) variety of Czech. During their pregnancy, none of the speakers lived abroad and all estimated their average daily production of Czech at 80 to 100%. All but one reported to talk to their unborn child every day (average 20 mins, range 5–60 minutes). None had hearing or speech disorders.

2.2. Recording procedure

The participants were recorded during spontaneous description of a set of pictures from a children's book; they were asked to describe any objects and activities displayed. Pictures were carefully chosen so that their content would elicit at least two instances of each of the 10 Czech monophthongs /ɪ iː ɛ ɛː a aː o oː u uː/. Recording took place in a sound-treated booth. The series of 6 target pictures were displayed on a computer screen in front of the participant who switched from one to the next picture at her own pace (prior to recording, speakers practiced describing a picture not contained in the recorded series). Participants described the same set of 6 pictures twice, once in IDS and once in ADS; with order counterbalanced across subjects (unusually talkative speakers were interrupted after the 3rd or 4th picture). For the elicitation of IDS, the speaker was in the booth on her own and was instructed to describe the pictures to her unborn child. For ADS, the speaker described the material to a female experimenter sitting beside her, facing the computer screen, and not interacting. Between the two recordings, there was a 10-minute break. All speakers produced 10–20 minutes of each IDS and ADS.

Recordings were done with a head-mounted condenser microphone AKGC 520 L and an Edirol UA 25 sound card connected to a Macbook running Audacity, at 44.1kHz sampling frequency and 16-bit quantization.

2.3. Data annotation

The recorded data were annotated in Praat [5]. First, content words containing the vowels of interest in a word-initial stressed syllable were segmented and labelled. Only those words that appeared both in IDS and ADS, mostly under phrasal accent, were selected for vowel segmentation (the words could differ in suffixes that did not affect the stem vowel and the number of syllables). Speakers recorded on average 7 tokens per vowel category per style (between-speaker range 3.5–10.4). The beginning and end points of vowels were determined from the waveform as zero crossings of the first and last vocalic periods resembling in shape the periods in the vowels' central parts; at the same time the spectrogram was checked to contain visible formants (especially F2) throughout the entire vowel interval.

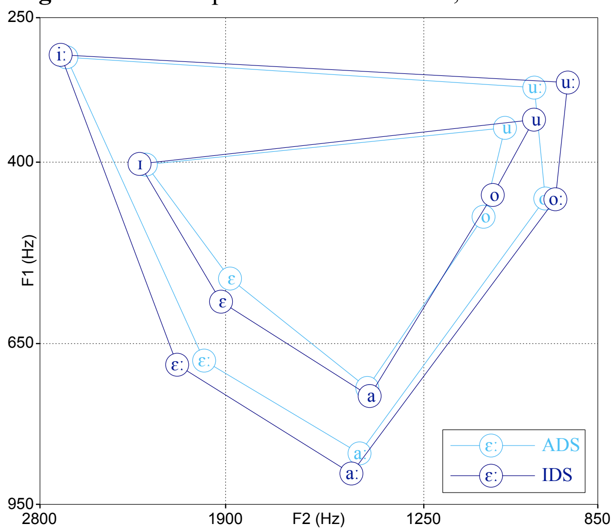
2.4. Acoustical analyses

Using Praat [5] the following "vowel properties" were analysed: F0, duration, F1 and F2. Duration was measured as the interval between the beginning and end points of a vowel; the values measured in seconds were log transformed. F0 was analyzed in the central 40% portions of the vowels using cross-correlation with the pitch range set to 120-400 Hz. If the initial analysis failed, F0 was reanalyzed with different settings, namely a lowered pitch floor, and if that failed, with lowered criterion for voicedness. The tokens for which even the third analysis failed were not further considered in the F0 measurements. Statistical analyses were done on F0 values in Mel.

Vowel formants were assessed over the central 40% portions of the vowels, using the Burg algorithm [1]. Formants were determined with the optimal ceiling method described in [13]. The optimal-ceiling method searches for such formant analysis settings that yield the lowest variation among each vowel category's measured formants. Here, within-speaker and within-style variation was optimized over the first three formants. Tokens for which the automated analysis yielded unlikely values (*n*=106) were reanalyzed manually. ERB-transformed F1 and F2 were further used to calculate the area of the F1-F2 vowel space per speaker and style. Vowel space areas were calculated separately for short and long vowels as the latter are consistently more peripheral and thus cover a larger space than their short counterparts (two speakers' /u:/ was missing: to calculate the area, its F1 and F2 were interpolated from /i:/ and /o:/).

VoiceSauce [31] was used to assess the speakers' "voice quality" in each speaking style. This analysis was done upon perceptual assessment of the data that revealed a noticeable degree of breathiness in some speakers' IDS. Parameters reflecting the vowels' harmonic organization, namely, cepstral peak prominence (CPP) and harmonic-to-noise ratio (HNR, in 0-3500 Hz), as well as intensity (measured in [5]) were explored and compared across styles.

Figure 1: Vowel spaces in IDS and ADS, ERB-scaled.



2.5. Statistical analyses

Vowel space area, F0, and duration were each submitted to three separate linear mixed effects (LME) models (*lmer* function in R, [2]). The models contained style and phonological length as fixed factors with orthogonal contrasts (ADS -.5 vs. IDS +.5, and long -.5 vs. short +.5). In the models for F0 and duration, phonological vowel quality was entered as another fixed factor with four orthogonal contrasts that tested each of the three corner vowel qualities against the two mid vowel qualities, and the two mid vowels against one another (*a* -.5 vs *e* +.25 and *o* +.25, *i* -.5 vs *e* +.25 and *o* +.25, *u* -.5 vs *e* +.25 and *o* +.25, *e* -.5 vs *o* +.5). Participant was entered as a random effect and random slopes were included for each of the within-subjects factors. These hypothesis-driven analyses were done with alpha .01 (Kenward-Roger approximation of *df*, package *pbkrtest* [20]).

The two measures of voice quality, CPP and HNR, were submitted to two separate exploratory LME models with style as the fixed factor (ADS -.5 vs. IDS +.5), and participant and vowel category as random effects, alpha was set to the less conservative .05.

3. RESULTS

Significant effects are listed in Table 1. Vowel space area is affected by vowel length: unsurprisingly, the space defined by short vowels has a smaller area than that of long vowels. Importantly, vowel space area is also affected by style: in prenatal IDS vowels cover a larger area than in ADS, as seen in Fig. 1.

Table 1: Significant effects in the vowel-properties models. Effects involving *style* are in bold.

model	effect	estimate	*t*-value
V-space	intercept	27.4	19.169
area	**style** (-ADS,+IDS)	**3.1**	**3.256**
(ERB^2)	length (-lng,+shrt)	-22.0	-15.952
F0	intercept	156	54.299
(Mel)	length (-lng,+shrt)	5	3.215
	-a vs. +eo	12	7.378
	-i vs. +eo	-7	-2.888
	-u vs. +eo	-13	-5.293
	style*length	**4**	**2.975**
	style*a vs. eo	**-9**	**-3.640**
	style*i vs. eo	**9**	**3.560**
	length*i vs. eo	12	4.976
duration	intercept	-2.345	-104.6
(ms)	**style** (-ADS,+IDS)	**0.107**	**4.832**
	length (-lng,+shrt)	-0.820	-36.758
	-a vs. +eo	-0.134	-4.250
	-i vs. +eo	0.364	6.889
	length*i vs. eo	-0.551	-9.653

F0 is affected by vowel length: short vowels have higher F0 than long ones. The two-way interaction of style and length and inspection of estimated means (*emmeans* package [24]) show that the short vs. long vowel difference in F0 is significant in IDS, and is smaller or non-existent in ADS; see Table 2. The effect of vowel quality on F0 reflects vowel-intrinsic F0: low vowels have lower F0 than mid vowels, and high vowels have higher F0 than mid vowels; inspection of the means suggests that the vowel-intrinsic F0 effect is stronger in ADS than in IDS.

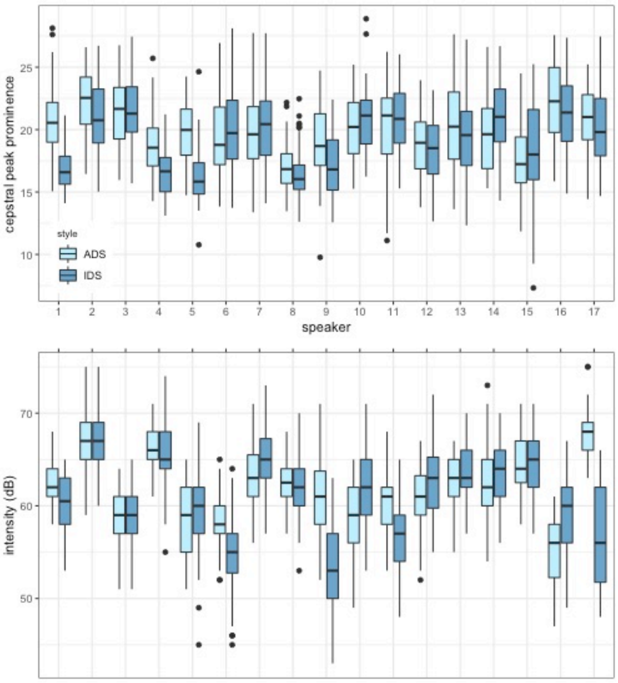
The main effect of phonological length on duration confirms that short vowels are acoustically shorter than long ones (the effect of length is smaller for the high front than for mid vowels, which is in line with previous findings on Czech /iː/-/ɪ/ [30]). The significant vowel quality contrasts confirm that, in line with vowel-intrinsic duration, low vowels are longer than mid vowels, and that the high front vowels are shorter than mid vowels. The effect of style reveals that vowels are longer in IDS than in ADS. The absence of a significant style by length interaction (est. -0.003, t = -0.099) suggests that the short-long durational differences are similar in ADS and IDS. A comparison of estimated means, given in Table 2, confirms that long vowels have longer duration than short vowels in both IDS and ADS, by a factor of 2.27 and 2.28, respectively.

Table 2: F0 and duration, means and 95% conf. int.

	ADS		**IDS**	
	short	long	short	long
F0 (Mel)	158	155	160	153
	151–165	148–161	153–166	146–159
dur (ms)	60	137	67	152
	58–63	130–144	63–71	141–165

As for voice quality, only the model for CPP yielded an effect of style (t = -2.102, p = .047) implying that CPP is higher in ADS than in IDS; see Fig. 2. A lack of significant intensity difference between ADS and IDS (est. -0.933, t = -1.072), along with the inspection of individual data in Fig. 2, indicate that the higher breathiness in IDS is not attributable (solely) to lower loudness.

Figure 2: CPP and intensity per speaker and style.



4. DISCUSSION

We examined whether a speech style similar to infant-directed speech occurs in utterances addressed to unborn children. Our data show that vowels in prenatal IDS cover a larger vowel space area than vowels in ADS, which is in line with a number of previous studies on postnatal IDS. Since the larger vowel space area in IDS is by definition caused by larger distances between the individual vowels, it could serve to facilitate speech sound processing, or even learning, already during fetal development.

In line with the 'postnatal' IDS literature, we found that vowels are longer in prenatal IDS with durational distinctions between phonologically short and long vowels being preserved, although not enhanced. An additional analysis of word durations and vowel/word duration ratios showed that the lengthening in prenatal IDS is due to a slower speech tempo: words were on average 0.052 seconds longer in IDS than ADS (t = 2.115), while no effect of style was detected for vowel/word ratios. Note that the slower speech tempo may also provide an alternative explanation for the enlarged vowel space: careful (i.e. slower) speaking style usually leads to more prototypical (i.e. peripheral) realization of vowels [25] that cover a larger area than less peripheral ones.

Although we did not detect any main effects on F0, we found that F0 may serve (along with duration) as a distinguishing cue between phonologically short and long vowels in prenatal IDS but not (or to a smaller extent) in ADS. This finding speaks for the development-facilitating function of prenatal IDS a bit more unambiguously than the finding for vowel space area. If F0 changes were emotional only, one would expect a higher (~affective) or a lower (~calming) F0 overall.[2]

The voice quality data speak in favour of the affective function of prenatal IDS. As shown by the lower cepstral peak prominence, prenatal IDS is breathier than ADS, which may be reflecting emotional attachment or intimacy. This finding aligns well with previous reports of higher breathiness in postnatal IDS and talkers' tendencies to employ a calming voice with the youngest infants [21, 28].

In conclusion, pregnant women address their unborn children in a style that one may call prenatal IDS. As in the debate on 'postnatal' IDS, it is as yet unclear to what extent the acoustic properties of prenatal IDS are aimed at helping the fetus learn the language and to what extent they are a by-product of the mother's emotional attachment.

[1] Funding from Charles University and a Czech Science foundation grant 18-01799S to KC. RS was supported by the European Regional Development Fund-Project "Creativity and Adaptability as Conditions of the Success of Europe in an Interrelated World" (CZ.02.1.01/ 0.0/0.0/ 16_019/ 0000734).

[2] The women reported to talk to their unborn child for ~20 minutes/day, which means that if at all aimed at facilitating language development, prenatal IDS is unlikely to be the main source of language input for the fetus.

5. REFERENCES

[1] Anderson, N. 1978. On the calculation of filter coefficients for maximum entropy spectral analysis. *Childres: Modern Spectrum Analysis*. IEEE, 252-5.
[2] Bates, D., Maechler, M., Bolker, B., Walker, S. 2015. Fitting linear mixed-effects models using lm4. *J Stat Softw*, 67, 1–48.
[3] Benders, T. 2013. Mommy is only happy! Dutch mothers' realisation of speech sounds in infant-directed speech expresses emotions, not didactic intent. *Infant Behav Dev* 36, 847–62.
[4] Bernstein Ratner, N., Luberoff, A. 1984. Cues to post-vocalic voicing in mother-child speech. *J Phon* 12, 285–9.
[5] Boersma, P., Weenink, D. 2018. Praat: doing phonetics by computer. Version 6.0.40, www.praat.org
[6] Burnham, D., Kitamura, C., Vollmer-Conna, U. 2002. What's new, pussycat? On talking to babies and animals. *Science* 296, 1435.
[7] Cooper, R. P., Aslin, R. N. 1990. Preference for infant-directed speech in the first month after birth. *Child Dev* 61(5), 1584–95.
[8] Cristia, A. 2010. Phonetic enhancement of sibilants in infant-directed speech. *JASA* 128, 424–34.
[9] Cristia, A. 2013. Input to language: The phonetics and perception of infant directed speech. *Lang Linguist Compass* 7/3, 157–70.
[10] DeCasper, A.J., Fifer, W.P. 1980. Of human bonding: Newborns prefer their mother's voices. *Science* 208, 1174–1176.
[11] Englund, K., Behne, D. M. 2005. Infant directed speech in natural interaction – Norwegian vowel quality and quantity. *J Psycholinguist Res* 34, 259–280.
[12] Englund, K. Behne, D. 2006. Changes in infant directed speech in the first six months. *Infant Child Dev* 15, 139–160.
[13] Escudero, P., Boersma, P., Rauber, A.S., Bion, R.A.H. 2009. A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *JASA* 126, 1379–93.
[14] Fernald, A., Simon, T. 1984. Expanded intonation contours in mothers's speech to newborns. *Dev Psych* 20, 104–13.
[15] Fernald, A., Taeschner, T., Dunn, J., Papousek, M. 1989. A cross-language study of prosodic modification in mothers' and fathers' speech to preverbal infants. *J Child Lang* 19, 477–501.
[16] Fish, M.S., García-Sierra, A., Ramírez-Esparza, N., Kuhl, P. 2017. Infant-directed speech in English and Spanish: Assessments of monolingual and bilingual caregiver VOT. *J Phon* 63, 19–34.
[17] Gervain, J. 2015. Plasticity in early language acquisition: the effects of prenatal and early childhood experience. *Curr Opinion Neurobiology* 35, 13–20.
[18] Graven, S.N, Browne, J.V. 2008. *Auditory Development in the Fetus and Infant*. Amsterdam: Elsevier.
[19] Green, J. R., Nip, I. S. B.,Wilson, E. M., Mefferd, A. S., Yunusova, Y. 2010. Lip movement exaggerations during infant-directed speech. *JSLHR* 53, 1529–42.
[20] Halekoh, U., Højsgaard, S. 2014. A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models. *J Stat Softw* 59, 1–30.
[21] Kitamura C., Thanavishuth, C., Burnham, D., Luksaneeyanawin, S. 2002. Universality and specificity in infant-directed speech: pitch modification as a function of infant age and sex in a tonal and non-tonal language. *Infant Beh Dev* 24, 372–92.
[22] Kuhl, P.K., Andruski, J.E., Chistovich, I.A., Chistovich, L.A., Kozhevnikova, E.V., et al. 1997. Cross-language analysis of phonetics units in language addressed to infants. *Science* 277, 684–686.
[23] Kuhl, P.K. 2004. Early language acquisition: cracking the speech code. *Nature Rev Neurosci* 5, 831–43.
[24] Lenth, R., Singmann, H., Love, J., Buerkner, P., Herve, M. 2018. R package *emmeans*, https://cran.r-project.org/web/packages/emmeans/ index.html
[25] Lindblom, B. (1963). Spectrographic study of vowel reduction. *JASA* 35(11), 1773–81.
[26] Liu, H.M., Kuhl, P.K., Tsao, F.M. 2003. An association between mothers' speech clarity and infants' speech discrimination skills. *Dev Sci* 6(3), 1–10.
[27] McMurray, B., Kovack-Lesh, K.A., Goodwin, D., McEchron, W. 2013. Infant directed speech and the development of speech perception: Enhancing development or an unintended consequence? *Cognition* 129, 362–78.
[28] Miyzawa, K., Shinya, T., Martin, A., Kikuchi, H., Mazuka, R. 2017. Vowels in infant-directed speech: More breathy and more variable, but not clearer. *Cognition* 166, 84–93.
[29] Moon, C., Cooper, R.P., Fifer, W.P. 1993. Two-day-olds prefer their native language. *Infant Beh Dev* 16, 495–500.
[30] Podlipský, V.J., Skarnitzl, R., Volín, J. 2009. High front vowels in Czech: A contrast in quantity or quality? *Proc Interspeech* Brighton, 132–5.
[31] Shue, Y., Keating, P., Vicenik, C., Yu, K. 2011. VoiceSauce: A program for voice analysis. *Proc 17th ICPhS* Hong Kong, 1846–49.
[32] Sundberg, U. 1999. Quantity in infant-directed speech. *Proc 14th ICPhS* San Francisco, 2189-91.
[33] Zhao, T., Moon, C., Lagercrantz, H., Kuhl, P.K. 2011. Prenatal motherese? Newborn speech perception may be enhanced by having a young sibling. *Psi Chi J Undergrad Res* 16, 90–94.