POLISH INFANT DIRECTED VS. ADULT DIRECTED SPEECH: SELECTED ACOUSTIC-PHONETIC DIFFERENCES

Agnieszka Czoska[1], Katarzyna Klessa[2], Maciej Karpiński[2]

[1]The Institute of Psychology, [2]The Institute of Linguistics, [1,2]Adam Mickiewicz University in Poznań, Poland
agaczoska@gmail.com, klessa@amu.edu.pl, maciej.karpinski@amu.edu.pl

ABSTRACT

Infant-directed speech (IDS) is reported to differ significantly from adult-directed speech (ADS) in its acoustic-phonetic properties. In IDS, phonetic features of individual speech sounds tend to be intensified [6, 14, 20]. An example phenomenon documented for IDS in several languages is vowel hyperarticulation [25]. Thus, the formant frequency values (F1, F2) vary in the two target speaking styles. Other modifications can be observed in F_0 levels (e.g., [5, 11]. Due to infants' preference towards IDS [5], laboratory-elicited IDS recordings are often used as stimuli in infant speech perception studies, aiming for example at the investigation of the effects of short-term exposure to foreign-language stimuli in early infancy and its potential contribution to the development of language learning skills (e.g., [15]). In the present study, we compare F_0, F1, F2 values, and segmental duration in vowels produced by five female speakers of Polish, reading pseudo-word lists in IDS and ADS.

Keywords: infant directed speech, adult directed speech, timing variability, formant differences.

1. INTRODUCTION

A common practice in most studies on speech perception by infants is to use stimuli recorded in infant-directed speech (IDS) because infants are assumed to differentiate between IDS and adult-directed speech (ADS), and to prefer the former one [5, 7, 24]. Preference towards IDS was confirmed for young preverbal infants, even neonates. Evidence for the preference in infants older than 9 months is less clear [20], but IDS is used in research on speech perception even in infants older than 12 months. Many possible explanations of such preference can be found in the literature: from ones pointing out that IDS facilitates acquisition of relevant phonetic features [6, 25], to those indicating that certain characteristics of IDS emotional content may obscure phonetic information [21]. It has been reported that in IDS, emotional information is in the foreground and proper emotional prosody may be favored over phonetic clarity since emotional prosody may distort phonetic information [18, 21]. Usually, the IDS stimuli used in infant speech perception studies are obtained by laboratory methods, either by reading or other type of elicitation techniques [2, 5], produced without physical presence of an infant addressee (but see [25]), or by using (semi-)synthetic speech prepared accordingly to findings reported on spontaneous IDS [15].

IDS differs from ADS in many respects (for a wider review see [20]). The difference seems to be intuitively known to naive language users. When instructed to speak or read “as if talking to an infant”, they produce speech different from what they produce under instruction “speak naturally” or “as if talking to an adult.” In spontaneous communication, speakers tend to adapt lexical and syntactical contents of utterances while for laboratory-elicited read speech, the most commonly modified properties belong to the phonetic-acoustic domain. A phenomenon characteristic for IDS across many languages is hyperarticulation of selected speech sounds and intensification of emotional prosody, especially in expressing positive emotions [18, 24]. Such intensification of features was reported for a number of languages, including English, Swedish, Finnish, Russian or Japanese [16, 14, 25].

As reported in many studies [14, 20, 6], spontaneous IDS is characterised by features related to “hyperarticulation” or “intensification” of vowels, including:

- vowel lengthening [20];
- “stretching of vowel space” [14], which means greater separation of different categories of vowels on F1 and F2 axes (or other continuous axis e. g. duration, cf. [25].

In the present study, F_0, formant (F1, F2) values, and segmental duration are compared for vowels produced by five female speakers of Polish reading pseudo-word lists in IDS, and in ADS.

2. SPEECH MATERIAL

2.1. IDS & ADS Corpus

The present study material belongs to a corpus of laboratory-elicited speech data from the following languages: Polish, French, Hungarian, Spanish, and Korean (further extensions are considered). The corpus has been developed by the present authors for the needs of the analysis of the development of the phonemic hearing and working memory in Polish infants using both behavioral methods, and electroencephalography and eyetracking-based techniques [3]. The assumption underlying the structure of the recording scenarios was to create sets of stimuli including various phoneme contrasts in IDS and ADS, with a special focus on contrasts non-existing or difficult in the infant's native language (Polish). The contrasts are based on, e.g., vowel quality (e.g., French vowel contrast for /e/ vs. /ɛ/ vs. /ø/)), vowel quantity (e.g., Hungarian pairs of long and short vowels /u/ vs. /u:/ or /ɑ/ vs. /a:/), the presence of aspiration in consonants (e.g., Korean /pʰ/ vs. /p/ or /kʰ/ vs. /k/). The realizations of the target phonemes were recorded using two types of lists designed for each contrast: a syllable list, and a two-syllable pseudo-word list. The syllable and word structures were designed to elicit realizations of the target sounds for each contrast in the same preceding and following contexts. In the nonsense-words, they were usually located in word-initial syllables but preferably not as the word-initial sounds.

Each speaker was requested to read each list twice, first in ADS and then in IDS. Before the second reading, they were asked to imagine a situation of speaking to an infant or to try saying a couple of sentences in IDS, and then to produce some of the pseudo-words they could remember using the same way of speaking. At the time of recording the list in the IDS, a picture of an infant was presented on the screen along with the stimuli.

The recordings took place in an anechoic chamber using a large membrane condenser microphone and an analogue-digital converter working at 44.1kHz sampling frequency and 16-bit resolution. The participants were asked to read a list of isolated pseudo-words. The stimuli were presented in random order with E-Prime software, in black letters against white background.

2.2. The present material and speakers

Before performing empirical studies with the participation of infants and small children, we regard it as necessary to conduct a series of experiments aimed at establishing a solid grounds for the infant studies by means of behavioural experiments with adult subjects (e.g., perception tests, see [3]) as well as by thorough phonetic-acoustic analyses of the recorded stimuli. The present work is a pilot study based on the Polish data only, given priority as the native language of the target group of listeners. Moreover, Polish data will be used as reference stimuli within the analyses based on data for other languages. A cross-language comparison will follow as a future work.

Thus, in order to compare selected acoustic-phonetic features in Polish IDS and ADS in the present study, we analyse pseudowords produced by five female speakers aged 24-36 (Mean = 28.2). In its present form, the IDS & ADS corpus includes only the recordings of female voices. All the selected pseudowords were composed of two CV syllables, such that all the analysed vowels were realised in the same consonantal contexts (the first syllable always began with a voiceless fricative /f/, while the second one started with /f/ or a voiceless dental stop /t/).

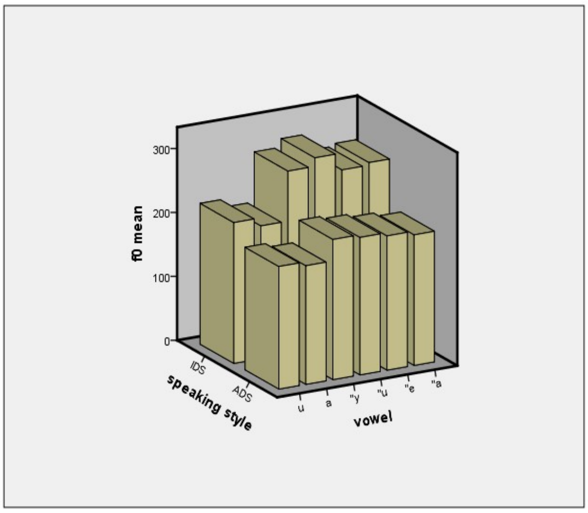
3. F_0, FORMANTS AND DURATIONS IN ADS AND IDS

3.1. Pitch frequency

Pitch frequency was measured in the realizations of pseudo-words using auto-correlation-based algorithm in Praat [4] for both ADS and IDS speaking styles for each speaker. Only measures taken within vowels were included in further analyses. We looked for differences in the pitch frequency values and pitch range between the speaking styles but we also controlled differences among the speakers.

Fig. 1 shows mean values of F_0 measured for vowels produced in stressed (initial) and unstressed (final) syllables. SAMPA was used for the corpus annotation, hence all labels referring to our material are based on the Polish SAMPA [23]. The difference in pitch between stressed and unstressed vowels appeared to be significant both in ADS and IDS but the effect is considerably stronger in IDS. One-way Anova calculated with vowel category as a grouping variable shows significant differences in pitch for vowels both in ADS and IDS. Further t-tests indicate that the most significant difference in pitch is found between vowels in stressed and unstressed positions (ADS: $F=4.89$, $p=0.027$; IDS: $F=174.26$, $p<0.001$). Significant differences have been also detected between ADS and IDS in the mean F_0 values in the initial (stressed) syllables ($F=1236.08$; $p<0.001$, $t=-28.345$; $df=962$), with the difference between the means of 73.92. Similarly, a significant difference was found between ADS and IDS in the mean F_0 values in the final syllable (i.e., second, unstressed syllable; F=91.561, p<0.001, t= – 9.978) with the difference in the means of 25.46. Accordingly, both the mean pitch values and their ranges were significantly higher in IDS, both for stressed and unstressed vowels. These results stay in line with the characteristics of IDS provided for other languages [20].

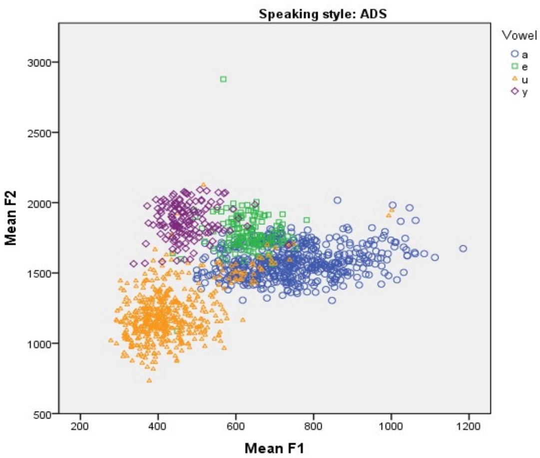
Figure 1: F_0 means in IDS and ADS: in stressed (/"e/, /"a/, /"y/, /"u/) unstressed (/u/, /a/) vowels



3.2. Formant frequencies

First (F1) and second (F2) formant frequencies were measured in the realisations of pseudowords using Praat [4] and Annotation Pro [12] in both ADS and IDS for each speaker.

Figure 2. Vowel space in ADS.



Automatic formant extraction may result in errors. For example, in speech with high pitch frequency the second and third formant (F2, F3) frequencies may be extracted instead of F1 and F2. The data from Praat analysis were manually corrected (85 instances out of 2655 (0.3%), 74 of them in IDS condition): by dividing values by 2. Here, we present the results of analyses performed on the corrected data.

Vowels in unstressed and stressed positions were analysed together. Figures 2 and 3 present vowel spaces in ADS and IDS speaking styles separately. In order to find whether those spaces differ significantly [6], data mining with machine learning algorithms k-Means and EM was performed using Weka 3.6 software [9]. Both k-Means and EM are clustering algorithms, EM was used also in [6]. In each case, 10-fold cross-validation was used as training method, and Euclidean distance as distance function in k-Means. The number of clusters was set manually for k-Means (4), while in the case of EM, it was found automatically in the process of model building. In ADS condition, EM algorythm found correctly four clusters, while in IDS five clusters were found (the performance over four clusters was also evaluated).

Figure 3: Vowel space in IDS

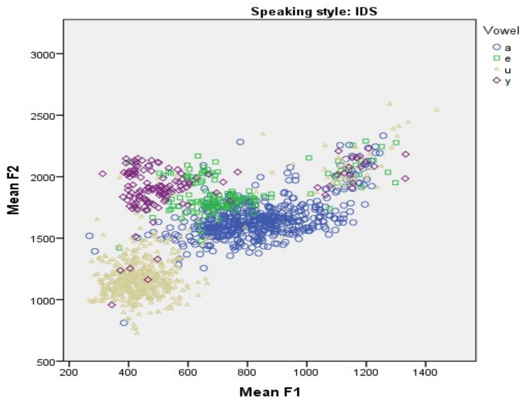


Table 1: Percent of correctly classified instances in data mining. EM model performance on IDS data is presented for both 5* and 4** clusters.

Clustering algorithm	ADS	IDS
k-Means	65.1%	76.7%
EM	81.7%	69.3 %*
		81.7%**

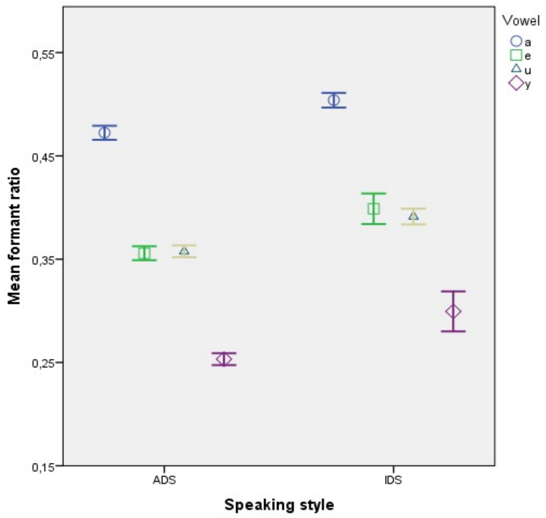
For further analyses, the difference between ADS and IDS formant ratio (F1/F2; FR) [10, 8] was calculated. One-way ANOVA showed significant differences in FR between vowels (F=691.9, p<0.001) and speaking styles (F=106.23, p<0.001), but not for the interaction of the factors (F=0.915, p=0.433) (see: Table 2 and Fig. 4). Post hoc (Tukey test) showed significant differences in FR between all vowel categories (within bot speaking styles) except /e/ and /u/ (p=0.958).

Table 2: Mean FR values in ADS and IDS speaking styles.

Vowel category (SAMPA labels)	ADS	IDS
/a/	0.472	0.504
/e/	0.356	0.399
/u/	0.356	0.391
/y/	0.253	0.276

The differences between ADS and IDS speaking styles are systematic and similar for all vowel categories. FR is higher in IDS than in ADS. The infant-directed speaking style results also in an increase of the number of outliers. We confirmed the observations reported in [6] that machine learning outcomes are better with IDS than ADS vowels (although with different algorithm was used for the present analyses).

Figure 4: Mean formant ratio (FR) in ADS and IDS.



3.2. Segmental durations

Durations of the four vowels in stressed positions (initial syllables) were extracted automatically from annotations using Annotation Pro [12] and compared between the two speaking styles. It was observed that IDS operates with significantly higher durations which may be partially caused by time requirements for more radical pitch changes. Differences in the mean duration of vowel realisations in stressed positions in IDS and ADS were confirmed by t-tests: /"a/ (F=39.93; p<0.001), /"e/ (F=6.65; p=0.010), /"u/ (F=57.76; p<0.001), /"y/ (F=12.58, p<0.001). Differences among mean durations of these vowels within speaking styles were not significant both in IDS (p=0.187) and ADS (p=0.052).

5. CONCLUSIONS & FUTURE WORK

Acoustic features of laboratory elicited IDS differ significantly and systematically from ADS utterances elicited under similar conditions. The present results are generally in line with earlier work reported for other languages (e.g. [17, 22]). Due to the need of further verification of the findings and because of the lack of ready-to-use reference material for Polish data, a necessary further task will be the design and analysis of a spontaneous IDS corpus acquired in natural environments for the needs of detailed comparisons with the laboratory elicited IDS.

The investigation reported in this paper belongs to a series of studies carried out in order to prepare solid and justified speech data for the needs of the analysis of infants' speech perception. One of the future applications of the results will be the improvement and extension of an IDS & ADS speech corpus currently developed by the authors for the needs of electroencephalography and eyetracking-based studies of the development of the phonemic hearing and working memory in infants.

6. ACKNOWLEDGEMENTS

This work is supported by the project: *NeuroPerKog: Development of the phonematic hearing and working memory in infants and children* funded by National Science Centre, Poland for years 2013-2018.

The authors also wish to thank the *NeuroPerKog* team at the **Centre for Modern Interdisciplinary Technologies at Nicolaus Copernicus University in Toruń**, Poland for their cooperation and advice in the process of corpus design.

7. REFERENCES

[1] Albin, D. D., Echols, C.H. 1996. Stressed and Word-final Syllables in Infant-directed Speech. *Infant Behariour and Development, 19*, 401-418.

[2] Albareda-Castellot, B., Pons, F., & Sebastián-Gallés, N. 2011. The acquisition of phonetic categories in bilingual infants: new data from an anticipatory eye movement paradigm. *Developmental science, 14*(2), 395-401.

[3] Klessa, K., Karpiński, M., Czoska, A. (in press) Design, structure, and preliminary analyses of a speech corpus of infant directed speech (IDS) and adult directed speech (ADS). Accepted at: *48th Annual Meeting of the Societas Linguistica Europaea*, Leiden.

[4] Boersma, P., Weenink, D. 2015. Praat: doing phonetics by computer (Version 5.1.05). [Computer program] Available: http://www.praat.org/

[5] Cooper, R. P., Aslin, R. N. 1990. Preference for infant-directed speech in the first month after birth. *Child development, 61*(5), 1584-1595.

[6] De Boer, B., Kuhl, P. K. 2003. Investigating the role of infant-directed speech with a computer model. *Acoustics Research Letters Online, 4*(4), 129-134.
[7] Englund, K. T. 2005. *Characteristics of and preference for infant directed speech.* Doctoral thesis, Faculty of Social Sciences and Technology Management, Norwegian University of Science and Technology, Trondheim.
[8] Hillenbrand, J., & Gayvert, R. T. (1993). Vowel classification based on fundamental frequency and formant frequencies. *Journal of Speech, Language, and Hearing Research*, *36*(4), 694-700.
[9] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I. H. 2009. The WEKA Data Mining Software: An Update; *SIGKDD Explorations*, *11*(1).
[10] Johnson, K. 2008. Speaker Normalization in Speech Perception. *The handbook of speech perception*, 363.
[11] Kitamura, C., Thanavishuth, C., Burnham, D., Luksaneeyanawin, S. 2001. Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant behavior and development, 24*(4), 372-392.
[12] Klessa, K., Karpiński, M., Wagner, A. 2013. Annotation Pro – a new software tool for annotation of linguistic and paralinguistic features. In D. Hirst & B. Bigi (Eds.) *Proceedings of the Tools and Resources for the Analysis of Speech Prosody (TRASP) Workshop, Aix en Provence*, 51-54.
[13] Koponen, E., Lacerda, F. 2003. Final lengthening in infant directed speech may function as a cue to phrase constituents. PHONUM, 9, 9-12.
[14] Kuhl, P. K., Andruski, 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.
[15] 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.
[16] Räsänen, O., Altosaar, T., & Laine, U. K. 2008. Comparison of prosodic features in Swedish and Finnish IDS/ADS speech. *Proc. of Nordic Prosody X*, Helsinki.
[17] Redford, M.A., Davis, B.L., Mikkulainen, R. 2004. Phonetic variability and prosodic structure in mothers. Infant Behavior and Development, 27, 477-498.
[18] Singh, L., Morgan, J. L., & Best, C. T., 2002. Infants' listening preferences: Baby talk or happy talk? *Infancy*, *3*(3), 365-394.
[19] Spence, M. J., Moore, D. S. 2003. Categorization of infant-directed speech: Development from 4 to 6 months. *Developmental psychobiology, 42*(1), 97-109.
[20] Soderstrom, M. 2007. Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review, 27*(4), 501-532.
[21] 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.
[22] Wang, Y., Seidl, A., Crista, A. (in press). Acoustic characteristics of infant-directed speech as a function of prosodic typology. In J. Heinz, R. Goedemans, H. van de Hulst. *Dimensions of Stress.* Cambridge, MA: Cambridge University Press.
[23] Wells, J. C. 1997. 'SAMPA computer readable phonetic alphabet'. 1997. In Gibbon, D., Moore, R., Winski, R. (Eds.). *Handbook of Standards and Resources for Spoken Language Systems.* Berlin and New York: Mouton de Gruyter. Part IV, section B.
[24] Werker, J. F., Pegg, J. E., McLeod, P. J. 1994. A cross-language investigation of infant preference for infant-directed communication. *Infant Behavior and Development, 17*(3), 323-333.
[25] Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., Amano, S. 2007. Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition, 103*(1), 147-162.