



MACIEJ KARPIŃSKI

Institute of Applied Linguistics, Faculty of Modern Languages and Literatures,  
Adam Mickiewicz University in Poznań

<https://orcid.org/0000-0001-5532-6188>

KATARZYNA KLESSA

Institute of Applied Linguistics, Faculty of Modern Languages and Literatures,  
Adam Mickiewicz University in Poznań

<https://orcid.org/0000-0002-0419-2295>

## The perception of non-native phonological categories in adult-directed and infant-directed speech: An experimental study

**ABSTRACT:** In the present study, we test whether adult listeners detect phonological contrasts faster and more accurately in non-native infant-directed speech (IDS) than in non-native adult-directed speech (ADS). 21 participants listened to pairs of speech signals and their task was to decide as quickly as possible whether the signals constitute the same or different words. Each pair of signals contained target vowels or consonants representing a certain category of contrast that was phonologically relevant in a given language but not in Polish, i.e., the native language of the participants of the listening test. The signals were presented in a random order, and each pair occurred in the material twice. Although we demonstrated significant acoustic-phonetic differences between the utterances realized in the IDS and ADS speaking styles, the listeners in our study were not significantly more accurate or faster in the identification of contrasts in either IDS or ADS stimuli.

**KEYWORDS:** infant-directed speech, native phonological categories, non-native phonological categories, L2 perception

Percepcja nienatywnych kategorii fonologicznych w mowie kierowanej do dorosłych i do dzieci. Badanie eksperymentalne

**STRESZCZENIE:** W badaniu przedstawionym w artykule podjęto próbę ustalenia, czy dorośli słuchacze trafniej i szybciej rozpoznają niewystępujące w ich rodzimym języku kontrasty fonologiczne w mowie skierowanej do niemowląt niż w mowie skierowanej do dorosłych. Grupę badaną stanowiło 21 osób. Uczestnicy odsłuchiwali pary sygnałów, a ich zadanie polegało na możliwie szybkim zadecydowaniu, czy sygnały są tymi samymi czy różnymi słowami. Każda z par sygnałów obejmowała kontrast oparty na samogłosce lub spółgłosce i relewantny w języku źródłowym, lecz nie w polskim, tj. rodzimym języku badanych. Boddźce prezentowano w losowej kolejności, a każdy z nich występował w materiale dwukrotnie. Badanie nie wykazało istotnej różnicy między trafnością i szybkością reakcji uczestników na boddźce oparte na mowie skierowanej do niemowląt oraz skierowanej do dorosłych, mimo iż wykazano istotne różnice fonetyczno-akustyczne w nagranych wypowiedziach dla obu stylów wypowiedzi.

**SŁOWA KLUCZOWE:** mowa kierowana do niemowląt, rodzime kategorie fonologiczne, obce kategorie fonologiczne, percepcja języka obcego

## Non-native speech perception and infant-directed speech (IDS) characteristics

Phenomena of non-native speech perception continue to be of interest to various research communities, also from the perspectives of L2 (second language) learning and teaching (Jamieson & Morosan, 1986; Flege et al., 1997; Best & Tyler, 2007), as well as speech disorders therapy or diagnostics (Kilman et al., 2015; Seery et al., 2013).

Non-native speech perception poses a range of challenges to learners who must acquire sensitivity to segmental and suprasegmental phenomena, including new sounds and phonological categories that may be absent from their L1 (first language) (Bradlow & Bent, 2008). Certain contrastive features absent from L1 inventory may be phonologically relevant in L2, while some of the contrastive features of L1 may turn to be irrelevant in L2 (Best et al., 2001; Best & Tyler, 2007; Polka, 1995). Even if learners are conscious of these issues, it often requires prolonged practice to acquire the ability to recognize new categories of contrasts in various natural settings of spontaneous communication (Kuhl et al., 2003; Cook et al., 2008; Lecumberri et al., 2010). For example, learning a language which features a two- or three-way vocalic duration contrast is difficult to those whose native languages do not involve contrasts in this dimension; similarly, native speakers of languages which feature duration-based contrasts, may find it difficult to abstract from them while using languages that do not employ phonological duration (Pajak & Levy, 2014). The ability to recognize and apply new phonological contrasts or abstract from some of the internalized ones remains crucial to understand and produce L2 utterances. Much time, attention and effort are often put in the process of learning these new contrasts both in research and practice (e.g., Jamieson & Morosan, 1986). Both teachers and students look for new means that would facilitate the task. One may hypothesize that some findings regarding the properties of infant-directed speech may suggest that it may have a facilitatory role in the acquisition and employment of these contrasts.

It has been reported that infant-directed speech (IDS, the speaking style typical of utterances addressed to infants and small children) contributes to better intersensory integration (Kitamura et al., 2014) and allows for more successful word learning in infants than adult-directed speech (ADS) (Zangl & Mills, 2007). IDS involves a range of facilitatory features that help infants to segment speech, to distinguish between speech sounds, and to acquire new phonological categories in L1 (e.g., Jusczyk et al., 1999; Thiessen & Saffran, 2005; Fernald & Simon, 1984; Trainor & Desjardins, 2002; Cristià, 2010; de Boer & Kuhl, 2003).

A general preference of infants towards maternal or infant-directed speech has been reported in many studies (e.g., Cooper et al., 1997). Speech prosody is

one of the domains for which the phenomenon has been confirmed: infants were observed to prefer utterances characteristic of IDS, i.e., featuring for example: higher fundamental frequency, higher formant ratios (F1/F2), slower speaking rates (e.g., Fernald et al., 1989; Czoska et al., 2015; Narayan & McDermott, 2016). On the segmental level, hyperarticulation is often mentioned as typical of IDS (Cristià & Seidl, 2014).

The properties of IDS and its role in speech development encourage researchers to explore its potential in L2 learning. This, however, poses a serious challenge because many other factors are involved and it is difficult to identify and isolate them in more comprehensive studies (Eaves et al., 2016).

## The aim of the study

As discussed above, much of the IDS characteristics may have emerged from the need of facilitating adult-infant communication, including speech perception enhancement in infants and young children. People communicating in cross-linguistic settings, including foreign language teachers, seem to intuitively employ some of the following features in their way of speaking: lower speech rate, more precise articulation, expressive intonation, in order to enhance speech intelligibility (Saito & van Poeteren, 2012). We suggest that systematic, extensive studies should be undertaken to inspect the relevance of IDS characteristics with regard to the L2 perception in various contexts such as teaching-learning situation, language acquisition, cross-cultural communication, or communication under adverse conditions, including speech or hearing disorders. Our research provides a contribution related to one of the most important aspects of speech perception, namely, to the ability of distinguishing phonological contrasts.

In the present study, we inspect whether adult listeners detect phonological contrasts in non-native IDS and in non-native ADS with similar precision and reaction times. The underlying research question is whether adult listeners exhibit faster and more accurate responses to IDS-based stimuli which would indicate the significance of facilitating properties of IDS. According to our knowledge, studies of this type and following the procedure employed here, have never been carried out with native speakers of any language, including Polish.

The stimuli set used in the present work consists of laboratory speech samples, i.e., isolated pseudowords recorded in two speaking styles: ADS and IDS. Even though contemporary research provides abundant evidence for the importance of using continuous speech data for studying communicative events as this helps to get closer to the reality of human communication, the case of infant directed

speech appears to be slightly different in this respect. The exposure to isolated words was found to be an important enhancing factor for early vocabulary development. Among others, a more frequent exposure of a child to isolated instances of a word may influence later knowledge of that word (Brent & Siskind, 2001).

## Speech perception test

In order to compare the correctness and reaction time to phonological contrasts in ADS and IDS for adult listeners, a perception test based on same-different discrimination procedure was designed. The task of the participants of the present study was to decide whether the signals presented as a pair (in a sequence) were the same or different words of an abstract language. The details of the testing procedure and scenario are further explained in this section.

## Participants of the study

The participants of the study were young Polish educated adults, students of philology, learning at least two foreign languages. The group included 16 females and 5 males, aged between 19 and 34 years. We selected individuals who did not declare any hearing or articulatory problems. All participants reported at least basic knowledge of English, three spoke French, nine reported to know some German.

## Speech material

Three languages were selected in order to provide representations of phonological contrasts that did not occur in Polish: German (DE), French (FR) and Korean (KR). In principle, the present study might have been based on completely abstract, artificial stimuli, not directly related to any native language. However, in order to increase ecological validity of the study, we decided to use contrasts existing in the phonological systems of natural languages, immerse them into phonotactically appropriate pseudowords, and have them read by native speakers. The target sounds of the French speakers' stimuli were front vowels: close-mid unrounded, open-mid unrounded, and close-mid rounded vowels produced word-internally in the pseudowords: /feta – feta – fɔta/. The German-based stimuli category contained contrastive fricatives: voiced uvular,

voiceless uvular, and voiceless glottal, produced word-initially in the pseudowords: / ata – χata – hata/. Finally, Korean-based pseudowords were based on stops: strongly aspirated, fortis unaspirated, and lenis (un)aspirated, produced word-internally in the pseudowords: /hap<sup>h</sup>a – hapa – haba/ (see also Table 1). Two-syllable pseudowords were used (vs. isolated phonetic segments) in order to present contrasts as parts of longer – presumably linguistic – units, and not just as individual sounds. With linguistic stimuli, participants should be more inclined towards the linguistic mode of categorical perception and differentiation, and not just to detailed acoustic comparison of sounds.

The speech signals were produced by each speaker in two speaking styles (infant-directed and adult-directed speech), in two separate sequences. In order to induce speakers to use infant-directed speaking style, images of infants were shown to them during IDS recording along with the written stimuli. Speakers were recorded in an anechoic chamber using a high-quality condenser microphone (Neumann TLM 103) and audio interface (Roland Studio Capture) working at 24bit / 44.1 kHz (cf. also. e.g., Czoska et al., 2015; Klessa et al., 2015).

Fundamental frequency and duration were measured in the stimuli in order to explore their general characteristics that would serve to verify whether the laboratory IDS shared key features of IDS indicated in literature as well as to provide a potential reference point for the interpretation of the perception test results. The measurements were taken with Annotation Pro (Klessa et al., 2013) using pYIN fundamental frequency estimator (Mauch & Dixon, 2014). Fundamental frequency was measured for all utterances. Only the values obtained for the vowels produced in the initial syllables of the pseudowords were included in the analysis. Final syllables were found to be problematic in terms of fundamental frequency extraction because of low energy and often irregular phonation.

As shown in Figure 1., the mean fundamental frequency value was systematically higher in IDS for all the three categories of stimuli (DE, FR, KR), with standard deviation of 50.6 for ADS and 68.2 for IDS. Syllable duration also differed between IDS and ADS. However, while it was higher in IDS for French- and Korean-based IDS pseudowords, the German-based ones were consistently shorter when produced as IDS (see Figure 2). Noteworthy, the phonetic-acoustic differences between ADS and IDS observed in the utterances produced by German and French speakers were more systematic than those produced by Koreans. Duration and pitch measurements are summarized in Table 1.

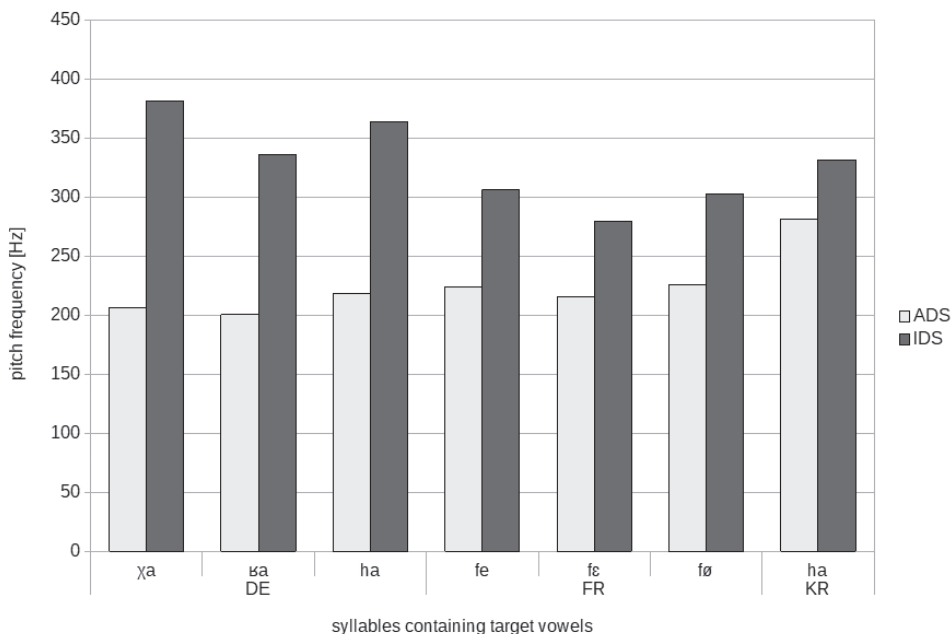


FIGURE 1. Fundamental frequency in ADS and IDS stimuli based on the three languages: German (DE), French (FR), Korean (KR). Measurements made on vowels of the initial syllables of respective signals.

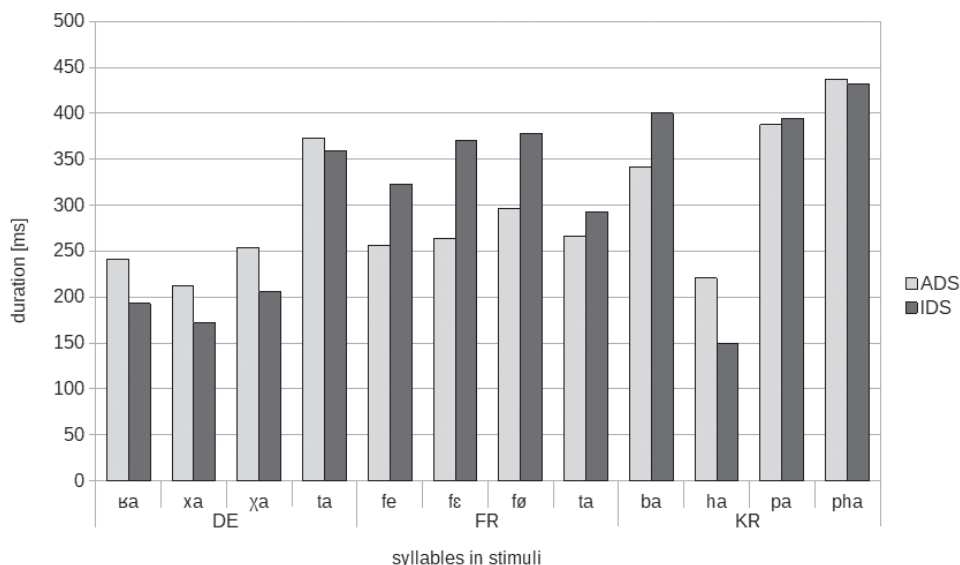


FIGURE 2. Mean syllable durations in ADS and IDS pseudowords based on German (DE), French (FR), and Korean (KR).

TABLE 1. Summary of the instrumental measurements of acoustic parameters

Stimulus category	Speaking style	Parameter	Mean	St. dev.
DE	ADS	duration (ms)	202.91	106.02
		pitch (Hz)	192.10	26.42
	IDS	duration (ms)	183.08	117.47
		pitch (Hz)	294.82	87.72
FR	ADS	duration (ms)	179.30	80.30
		pitch (Hz)	208.62	44.58
	IDS	duration (ms)	216.51	113.66
		pitch (Hz)	291.99	45.88
KR	ADS	duration (ms)	203.02	113.67
		pitch (Hz)	255.37	51.49
	IDS	duration (ms)	186.14	141.67
		pitch (Hz)	332.75	34.05

### The design of stimuli for the perception test

For each of the three categories of stimuli (DE, FR, and KR), we used three pseudowords in each of the two speaking styles (ADS and IDS), recorded by two female native speakers (3 languages  $\times$  3 pseudowords  $\times$  2 speaking styles  $\times$  2 speakers = 36). The design of perception test stimuli is shown in Table 2. Speech signals were paired to form 'same' or 'different' stimuli. Each stimulus, composed of two signals, occurred twice in the material, each time however, with a different order of pseudowords (3 languages  $\times$  2 speakers  $\times$  2 speaking styles  $\times$  6 pseudoword pairs  $\times$  2 occurrences = 144 two-pseudoword stimuli).

TABLE 2. Perception test stimuli design

Language	Contrast description	Stimuli pairs
Korean	Realizations of the pseudowords: /hap <sup>h</sup> a – hapa – haba/	Different: /hap <sup>h</sup> a – hapa/ /hap <sup>h</sup> a – haba/ /hapa – haba/ Same: /hap <sup>h</sup> a – hap <sup>h</sup> a/ /hapa – hapa/ /haba – haba/
	Target sounds – consonants produced word-internally: /p <sup>h</sup> – p – b/	

cd. table 1

German	Realizations of the pseudowords: /ʁata – χata – hata/ Target sounds – consonants produced word-initially: /ʁ – χ – h/	Different: /ʁata – χata/ /ʁata – hata/ /χata – hata/ Same: /χata – χata/ /ʁata – ʁata/ /hata – hata/
French	Realizations of the pseudowords: /feta – feta – fɔta/ Target sounds – vowels produced word-internally: /e – ε – ø/	Different: /feta – feta/ /feta – fɔta/ /feta – fɔta/ Same: /feta – feta/ /feta – feta/ /fɔta – fɔta/

### Experiment procedure

The stimuli prepared in the way explained above were presented to the group of participants of the study. An adequate script was prepared for Open Sesame software (ver. 3.1, Mathôt et al., 2012). The procedure comprised two stages: (1) the training stage in which participants observed how the procedure was implemented, learnt which keys to use for answers, and adjusted audio volume to a comfortable level; and (2) the experimental stage where the target stimuli were presented and measurements were taken. Each participant listened to 144 stimuli (pairs of pseudowords) via high quality semi-open headphones (AKG K240 Studio) in a quiet room at a comfortable volume individually adjusted during the training stage. The order of stimuli was randomly generated by the software. Each listener was informed that s/he would be presented pairs of words coming from a foreign language and was asked to decide as quickly as possible whether the 'words' within each pair were the same or not. The decision was communicated by pressing a key on the computer keyboard. His/her choice (yes/no) and reaction time (RT) were registered. On average, the entire procedure took ca. 20 minutes.



## Perception test results and findings

Perception test results were saved by OpenSesame to .CSV table format files and then imported to statistics software (IBM SPSS Statistics, ver. 26.0.0.0, IBM Corp. 2019). After the first data scrutiny, a number of extreme RT values were found that might have resulted from participants distraction or technical issues. In order to find the outliers, z-score normalization was carried out and the measurements with normalized RT values higher than 3 were rejected. Further testing for the significance of differences between means was based on the T-test and one-way ANOVA, as provided by the SPSS package.

The differences in the proportions of correct responses to ADS and IDS stimuli are illustrated in Figure 3, while Figure 4 displays bar plots of mean reaction times (RT, in milliseconds) to the contrasts represented by ADS vs. IDS stimuli pairs of three categories (German-, French-, and Korean-based). One can immediately see that Korean-based stimuli required significantly more time from the participants to respond (DE: mean = 698 ms SD = 616; FR: mean = 704 ms and SD = 560; KR: mean = 843 ms SD = 743.5;  $p < 0.01$ ,  $F = 16.252$ ), and that the proportion of correct responses was still lower than for the stimuli based on French or German (DE: 90%; FR: 91%; KR: 82%; the values in Figure 3 are presented separately for ADS and IDS). The participants successfully responded to more than 80% of stimuli of each category (source language and speaking style). For all the categories, highly significant differences in the mean RT values were found between responses to 'same' and 'different' stimuli: 'Same' stimuli required significantly shorter response times in all the stimuli categories (i.e., in DE, FR, KR;  $p = < 0.01$  in each case) and had a higher percentage of correct responses, as shown in Figures 8–10. Only for German-based stimuli within the 'same' category, a significant difference of RT between IDS and ADS was observed ( $p = 0.045$ ,  $F = 4.040$ ). For French-based and Korean-based stimuli, the difference was not significant even in the case of 'same' stimuli. Within the 'different' (i.e., containing a contrast) category, differences in the RT between contrasts were significant for German- and French-based stimuli but not for the Korean ones (DE  $p = 0.031$ , FR  $p = 0.003$ , KR  $p = 0.51$ ).

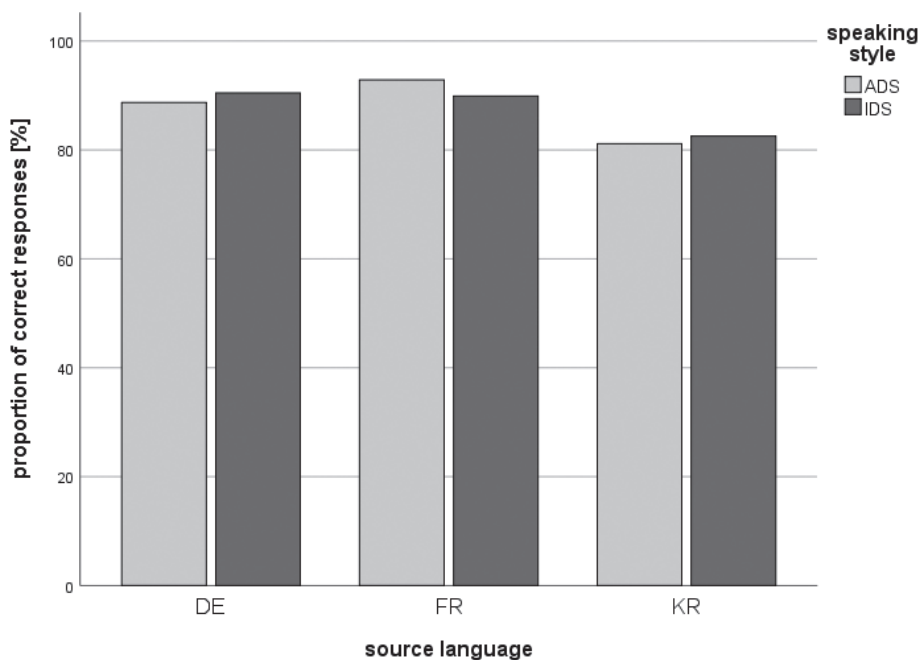
Figures 5, 6 and 7 show mean reaction times for all the types of stimuli. Stimuli based on the same pairs of signals but in different order are represented separately. In Figures 8–10, the same data are grouped by contrasts, showing reaction times jointly for the stimuli composed of the same signals in different order, and for all the 'same' stimuli as a separate category. The proportions of correct responses are presented in a similar way in Figures 11–13.

Korean /p – p<sup>h</sup>/ contrast appeared to be the most difficult for the Polish listeners (out of all the presented contrasts) as shown by the number of incorrect

answers as well as RT values. The /p – p<sup>h</sup>/ was correctly identified in only approximately 30 cases in ADS and 40 cases in IDS.

For the German-based stimuli, the contrast /χ – h/ proved to be more difficult than /ʁ – χ/ and /ʁ – h/ as shown by both the proportions of correct answers and RT values. Notably, the /χ – h/ contrast was slightly better recognized in IDS than in ADS (approx. 60 cases identified correctly in IDS vs. 40 correct cases in ADS).

The listeners appeared to be relatively good at the vowel contrasts based on French (approx. 80% for /ɛ – e/ and /ø – e/ and almost 100% correct answers for /ɛ – ø/). That might be surprising considering the fact that the listeners were native speakers of Polish, i.e., a consonantal language, with a relatively small vowel inventory.



a  
FIGURE 3. The proportion of correct responses (%) to ADS and IDS stimuli based on three languages (DE – German, FR – French, KR – Korean)

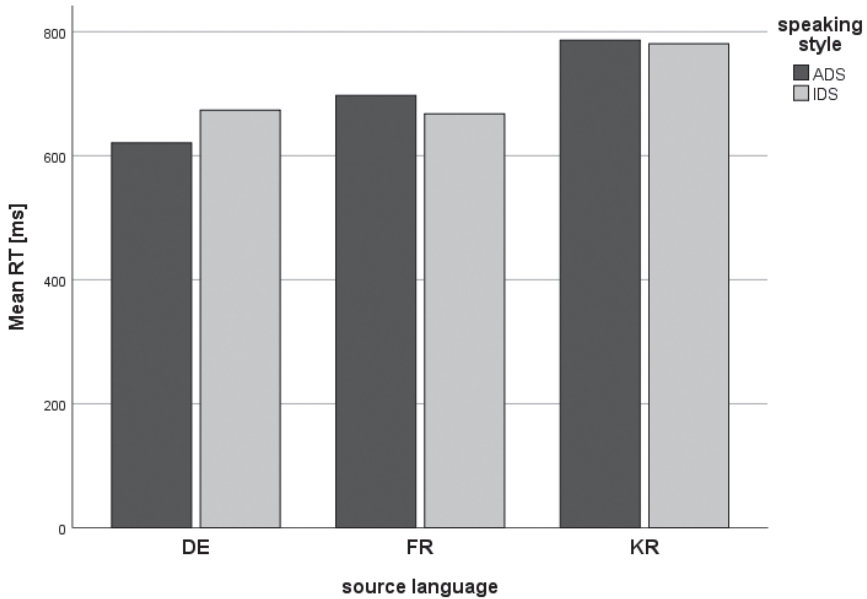


FIGURE 4. Same-different reaction times (RT, in milliseconds) for pairs of pseudowords based on three languages (German – DE, French – FR, and Korean – KR)

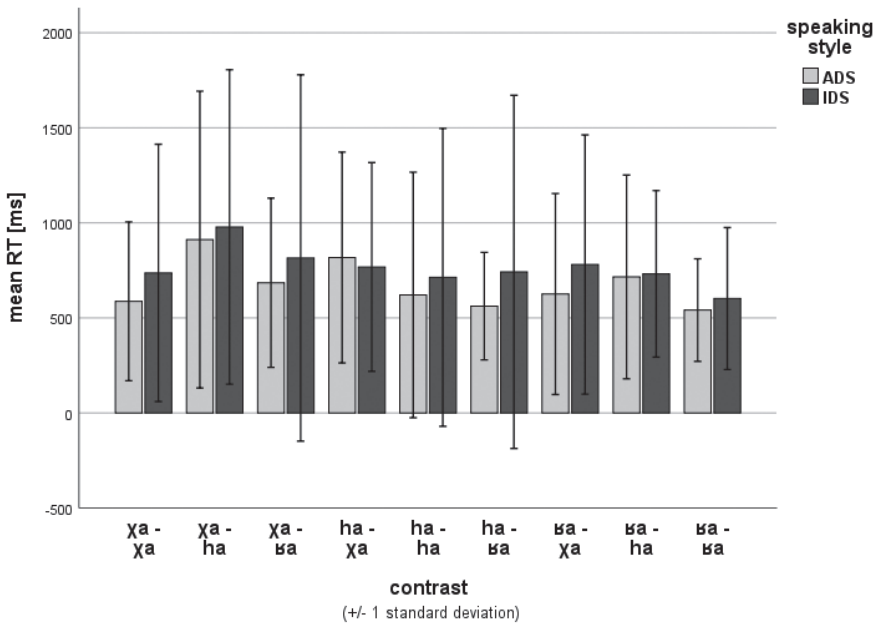


FIGURE 5. Mean reaction times (RT, in milliseconds) for German-based IDS and ADS stimuli. Labels indicate syllables in which contrasts were present in the pseudoword stimuli.

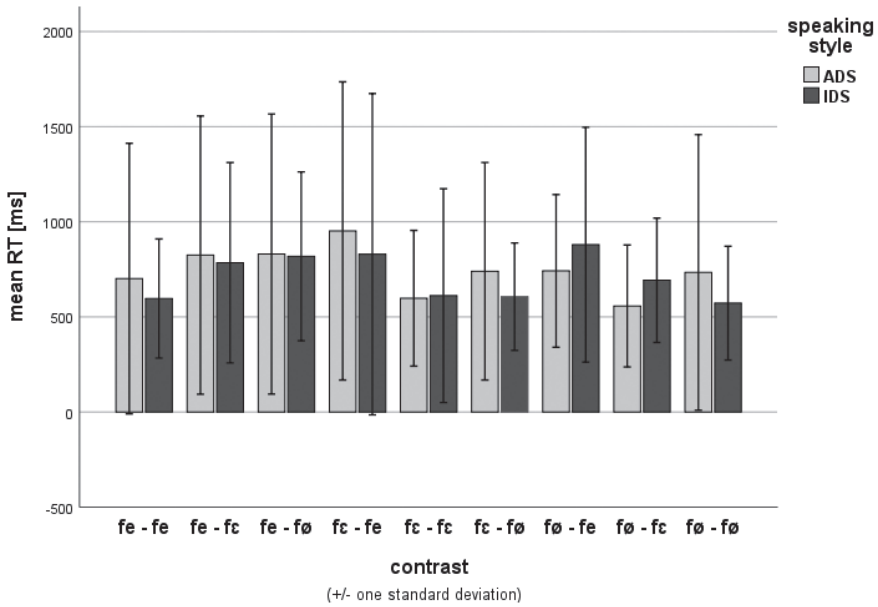


FIGURE 6. Mean reaction times (RT, in milliseconds) for French-based IDS and ADS stimuli. Labels indicate syllables in which contrasts were present in the pseudoword stimuli

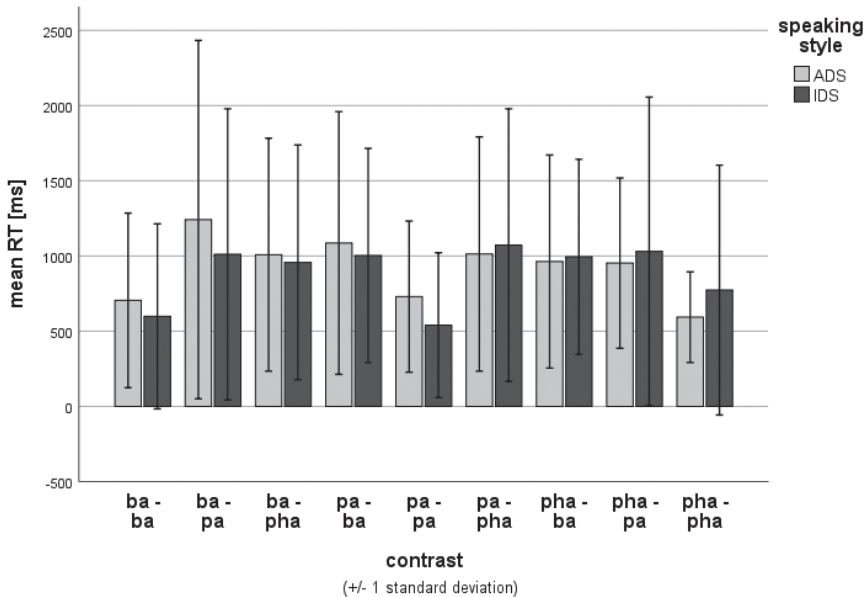


FIGURE 7. Mean reaction times (RT, in milliseconds) for Korean-based IDS and ADS stimuli. Labels indicate syllables in which contrasts were present in the pseudoword stimuli

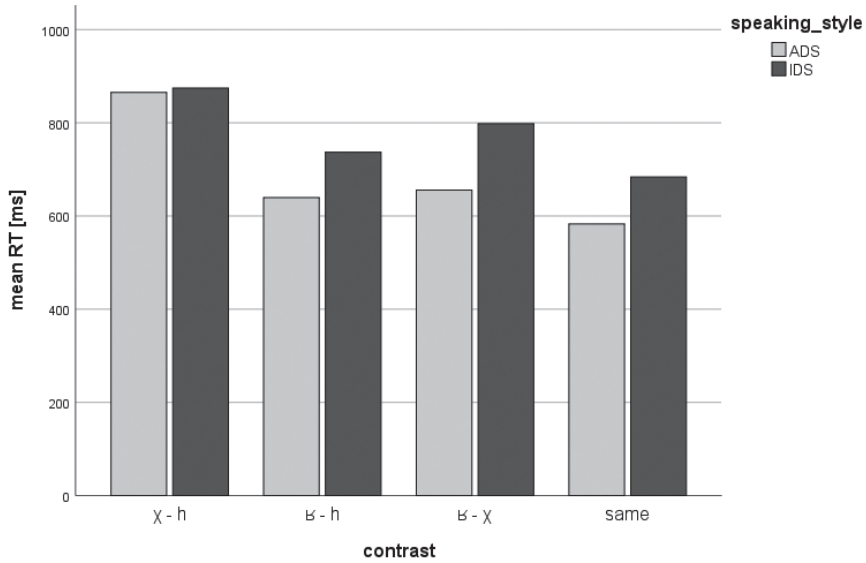


FIGURE 8. Mean reaction times (RT, in milliseconds) for German-based stimuli categorized by the contrast type (stimuli built of the same signals in different orders are treated jointly; RTs for 'same' are accumulated as one category)

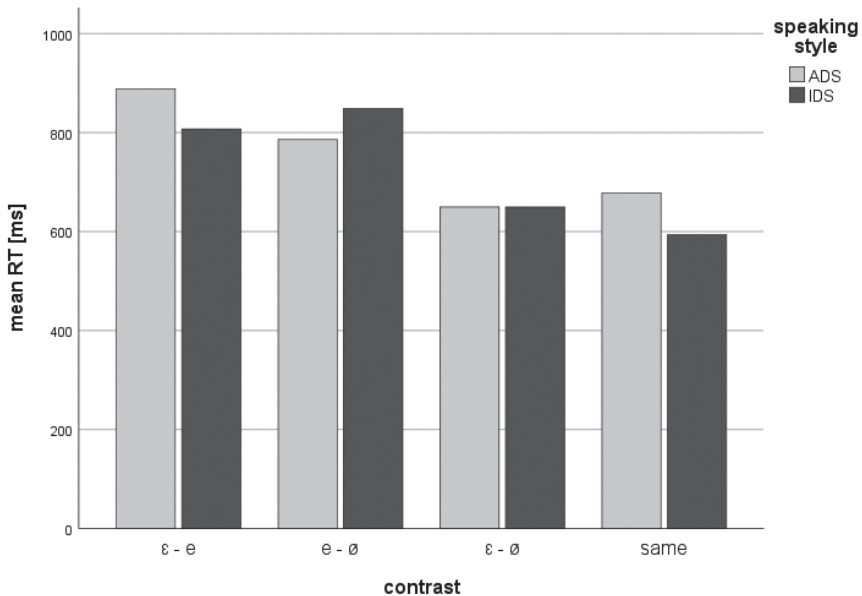


FIGURE 9. Mean reaction times (RT, in milliseconds) for French-based stimuli categorized by the contrast type (stimuli built of the same signals in different orders are treated jointly; RTs for 'same' are accumulated as one category)

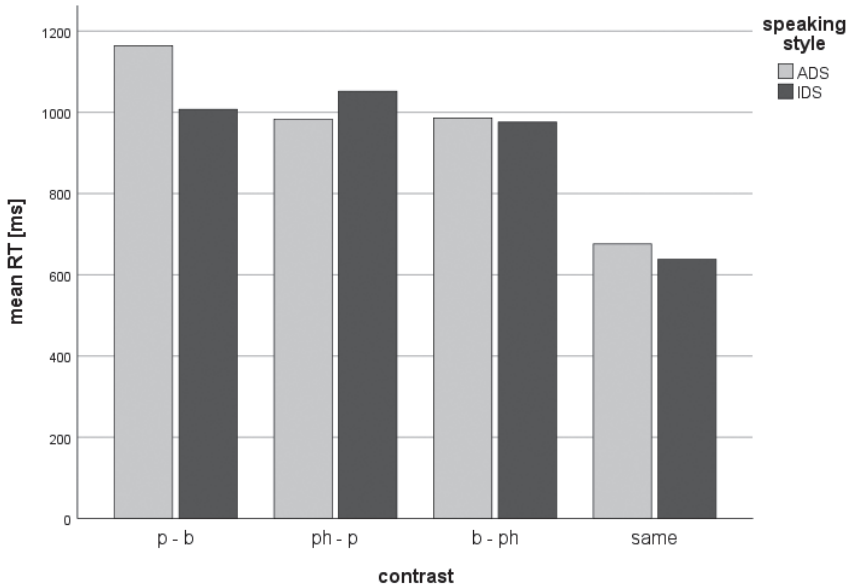


FIGURE 10. Mean reaction times (RT, in milliseconds) for Korean-based stimuli categorized by the contrast type (stimuli built of the same signals in different orders are treated jointly; RTs for 'same' are accumulated as one category)

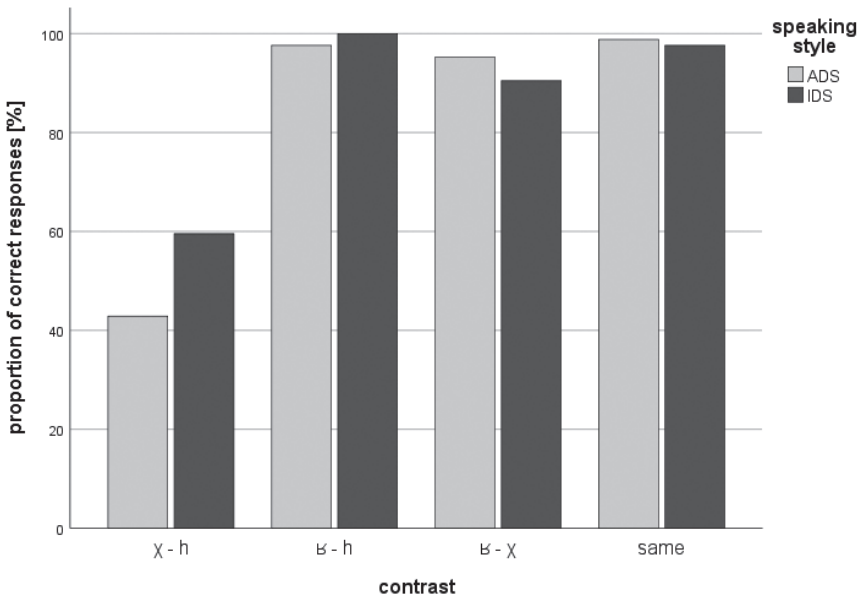


FIGURE 11. Proportions of correct responses (%) for the German-based stimuli categorized by the contrast type (stimuli built of the same signals in different orders are treated jointly; RTs for 'same' are accumulated as one category)

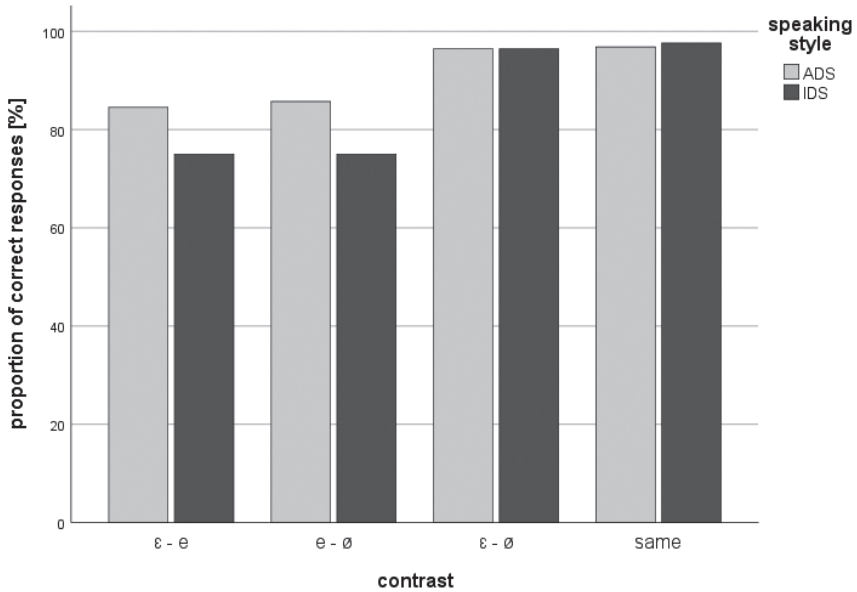


FIGURE 12. Proportions of correct responses (%) for French-based stimuli categorized by the contrast type (stimuli built of the same signals in different orders are treated jointly; RTs for 'same' are accumulated as one category)

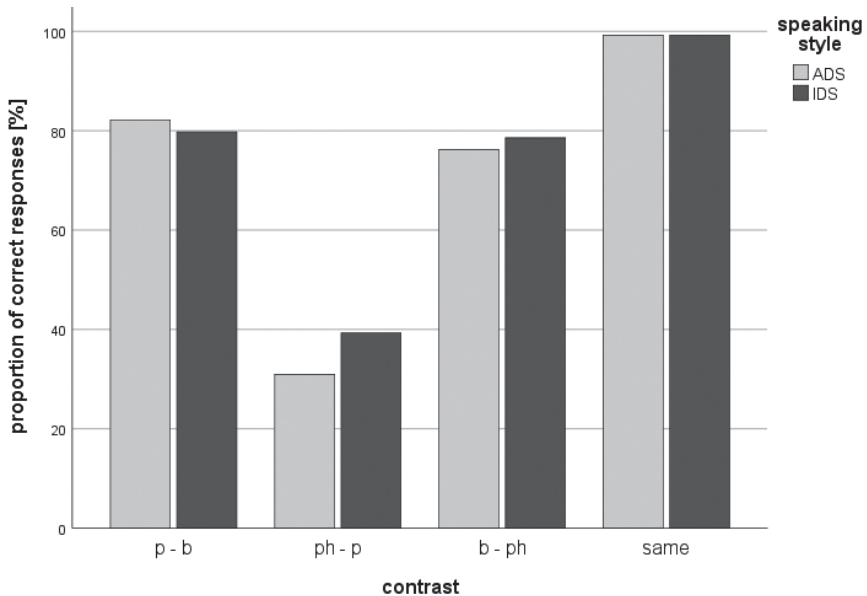


FIGURE 13. Proportions of correct responses (%) for Korean-based stimuli categorized by the contrast type (stimuli built of the same signals in different orders are treated jointly; RTs for 'same' are accumulated as one category)

## Conclusions and future work

As shown in a number of available studies, IDS provides a range of peculiar features that support speech segmentation as well as differentiation and identification of speech sounds (e.g., Jusczyk et al., 1999; Thiessen & Saffran 2005; Trainor & Desjardins, 2002; Cristià, 2010). Potentially, these features could also help adults to distinguish between phoneme categories in foreign languages. Our study was intended to test whether selected prosodic-acoustic features typical of infant-directed speech (IDS) facilitate the identification of foreign phonological contrasts by adult listeners.

Participants of the study listened to stimuli built based on contrasts found in three languages (German, French and Korean) but absent from their native language (Polish). Contrasts were presented in two-syllable pseudoword containers: French-based (FR): /feta – feta – fəta/, German-based (DE): /ʁata – χata – hata/, and Korean-based (KR): /hap<sup>h</sup>a – hapa – haba/. Pseudowords were produced by female native speakers of the respective source languages in two speaking styles: IDS and ADS. In order to find whether our laboratory IDS recordings share main prosodic features of naturally occurring IDS mentioned in the literature (e.g., Kuhl et al., 1997; Narayan & MacDermott, 2016; Adriaans & Swingley, 2017), the signals were analyzed and compared to their ADS counterparts. As expected, fundamental frequency and its range were significantly higher in IDS than in ADS for all the stimuli categories (DE, FR, KR). Significant differences were also present in the duration variability. However, German and French speakers tend to differentiate durations (of sounds, syllables, words) more systematically between IDS and ADS than Koreans.

The signals that met our technical and methodological requirements were used to build perception test stimuli. Each stimulus was based on a pair of signals that included both ‘different’ (with a contrast) and ‘same’ (no contrast) categories. Before the experiment, we explained, as a part of instruction for the participants, that the words came from an unknown foreign language. Then we asked the participants to decide as quickly as possible if the words they would hear in each pair were the same or different. We measured both the correctness of responses (same-different) and respective reaction times. In spite of the assumed potential of IDS support for contrast recognition (expressed by slower speech, hyperarticulation, exaggerated intonation) and some similarities between IDS and foreigner-directed speech (Uther et al., 2007), only a very limited influence of the speaking style (IDS vs. ADS) on phonological contrast detection was found in our listeners. Noteworthy, they tended to deal significantly better with ‘same’ than with ‘different’ stimuli both in terms of the number of correct responses and reaction times. But the differences in reaction



times between ADS and IDS stimuli for various contrasts were, in principle, not significant.

One must stress that the contrast detection task in the present study was based on isolated pseudowords and therefore it might have worked in a different way than in natural communicative settings. Although contrasts were placed in pseudowords in order to create the impression of linguistic stimuli, the pseudowords themselves were relatively short (two-syllable) and presented in isolation. Interactivity and other factors typical of parent-infant communication were absent from the laboratory conditions. Another potential problem is, as Eaves et al. suggest (2016), that the naturally observed effect of exposure to IDS may require a substantial number of stimuli occurrences, which would require to extend our experimental procedure with a learning or training stage. These and similar limitations are often inherent to laboratory studies, being a price to be paid for precise measurements of responses and for well-controlled stimuli.

The participants of the present experiment came from linguistic faculties which undoubtedly influenced their language skills. Therefore, a follow-up study is planned with other groups of listeners, including children and adults with no background in linguistics. Due to technical limitations, the study was focused on a small number of contrasts. In future studies, it might be valuable to produce a more complete picture of the phenomenon by covering a wider range of contrasts from different phonological dimensions.

Our materials, procedures, and software solutions will find applications in further inspection of ADS and IDS perception. The speech samples used as the basis for the perception test stimuli come from a larger corpus, designed for the purposes of investigation of the development of phonemic hearing and working memory in infants and children by means of electroencephalography, eye-tracking, and perception-based studies (Acknowledgement 1). The corpus provides rich resources, including other stimuli categories based on languages such as Spanish, Chinese or Hungarian, for our future studies. Further studies of IDS-specific facilitating mechanisms in non-native speech perception in different age groups, for different languages and settings, may help to revise the understanding of the Critical Period and revise some models of L2 adult learning or acquisition. As a consequence, it might possibly open new pathways for L2 phonetic and phonological training.

### Acknowledgements

The speech recordings, data processing as well as perception test design and implementation were supported by the grant: *NeuroPerKog: Development of the phonematic hearing and working memory in infants and children* funded by National Science Centre, Poland for years 2013–2019 (UMO-2013/08/W/HS6/00333).

The data mining and interpretation process was partly supported by the grant: COLING MSCA RISE Horizon 2020 Project: Minority Languages, Major Opportunities. Collaborative Research, Community Engagement and Innovative Educational Tools (Grant Agreement ID: 778384).

## 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), 3070–3078.
- BEST, C.T., McROBERTS, G.W., & GOODELL, E. (2001). Discrimination of Non-native Consonant Contrasts Varying in Perceptual Assimilation to the Listener's Native Phonological System. *The Journal of the Acoustical Society of America*, 109(2), 775–794.
- BEST, C.T., & TYLER, M.D. (2007). Nonnative and Second-language Speech Perception: Commonalities and Complementarities. In: O.-S. BOHN, M.J. MUNRO (eds.), *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, vol. 1334. John Benjamins Publishing Company.
- BRADLOW, A.R., & BENT, T. (2008). Perceptual Adaptation to Non-native Speech. *Cognition*, 106(2), 707–729.
- BOER, B. DE, & KUHL, P.K. (2003). Investigating the Role of Infant-directed Speech with a Computer Model. *Acoustic Research Letters Online*, 4, 129–134.
- BRENT, M.R., & SISKIND, J.M. (2001). The Role of Exposure to Isolated Words in Early Vocabulary Development. *Cognition*, 81, B33–B44. [https://doi.org/10.1016/S0010-0277\(01\)00122-6](https://doi.org/10.1016/S0010-0277(01)00122-6).
- CRISTIÀ, A. (2010). Phonetic Enhancement of Sibilants in Infant-directed Speech. *The Journal of the Acoustical Society of America*, 128(1), 424–434.
- CRISTIÀ, A. (2013). Input to Language: The Phonetics and Perception of Infant-directed Speech. *Language and Linguistics Compass*, 7, 157–170. <https://doi.org/10.1111/lnc3.12015>.
- CRISTIÀ, A., & SEIDL, A. (2014). The Hyperarticulation Hypothesis of Infant-directed Speech. *Journal of Child Language*, 41, 913–934. <https://doi.org/10.1017/S0305000912000669>.
- CZOSKA, A., KLESSA, K., KARPIŃSKI, M. (2015). Polish Infant Directed vs. Adult Directed Speech: Selected Acoustic-Phonetic Differences. *Proceedings of the 18<sup>th</sup> International Congress of Phonetic Sciences*. Glasgow, UK.
- COOKE, M., GARCIA LECUMBERRI, M.L., & BARKER, J. (2008). The Foreign Language Cocktail Party Problem: Energetic and Informational Masking Effects in Non-native Speech Perception. *The Journal of the Acoustical Society of America*, 123(1), 414–427.
- COOPER, R.P., ABRAHAM, J., BERMAN, S., & STASKA, M. (1997). The Development of Infants' Preference for Motherese. *Infant Behavioral Development*, 20, 477–488.
- EAVES, B.S., JR., FELDMAN, N.H., GRIFFITHS, T.L., & SHAFTO, P. (2016). Infant-directed Speech Is Consistent with Teaching. *Psychological Review*, 123, 758–771. <https://doi.org/10.1037/rev0000031>.
- FERNALD, A., & SIMON, T. (1984). Expanded Intonation Contours in Mothers' Speech to Newborns. *Developmental Psychology*, 20(1), Jan 1984, 104–113.
- FERNALD, A., TAESCHNER, T., DUNN, J., PAPOUŠEK, M., BOYSSON-BARDIES, B.D., & 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.

- FLEGE, J.E., BOHN, O.S., & JANG, S. (1997). Effects of Experience on Non-native Speakers' Production and Perception of English Vowels. *Journal of Phonetics*, 25(4), 437–470.
- IBM Corp. Released 2019. *IBM SPSS Statistics for Windows, Version 26.0*. Armonk, NY: IBM Corp.
- JAMIESON, D.G., & MOROSAN, D.E. (1986). Training Non-native Speech Contrasts in Adults: Acquisition of the English /ð/ – /θ/ Contrast by Francophones. *Perception & Psychophysics*, 40(4), 205–215.
- JUSCZYK, P.W., HOUSTON, D.M., & NEWSOME, M. (1999). The Beginnings of Word Segmentation in English-learning Infants. *Cognitive Psychology*, 39, 159–207.
- KILMAN, L., ZEKVELD, A., HÄLLGREN, M., & RÖNNBERG, J. (2015). Native and Non-native Speech Perception by Hearing-impaired Listeners in Noise-and Speech Maskers. *Trends in Hearing*, (19). <https://doi.org/10.1177/2331216515579127>.
- KITAMURA, C., GUELLAÏ, B., & KIM, J. (2014). Motherese by Eye and Ear: Infants Perceive Visual Prosody in Point-line Displays of Talking Heads. *PLOS ONE*, 9(10), e111467. <https://doi.org/10.1371/journal.pone.0111467>.
- KLESSA, K., KARPIŃSKI, M., & CZOSKA, A. (2015). Design, Structure, and Preliminary Analyses of a Speech Corpus of Infant Directed Speech (IDS) and Adult Directed Speech (ADS). Presented at 48<sup>th</sup> Annual Meeting of Societas Linguistica Europea (SLE). Leiden, The Netherlands.
- KUHL, P.K., et al. (1997). Cross-language Analysis of Phonetic Units in Language Addressed to Infants. *Science*, 277(5326), 684–686.
- 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.
- LECUMBERRI, M.L.G., COOKE, M., & CUTLER, A. (2010). Non-native Speech Perception in Adverse Conditions: A Review. *Speech Communication*, 52(11–12), 864–886.
- MATHÔT, S., SCHREIJ, D., & THEEUWES, J. (2012). OpenSesame: An Open-source, Graphical Experiment Builder for the Social Sciences. *Behavior Research Methods*, 44(2), 314–324. <https://doi.org/10.3758/s13428-011-0168-7>.
- MAUCH, M., & DIXON, S. (2014). pYIN: A Fundamental Frequency Estimator Using Probabilistic Threshold Distributions. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 659–663.
- NARAYAN, C.R., & McDERMOTT, L.C. (2016). Speech Rate and Pitch Characteristics of Infant-directed Speech: Longitudinal and Cross-linguistic Observations. *The Journal of the Acoustical Society of America*, 139(3), 1272–1281.
- PAJAK, B., & LEVY, R. (2014). The Role of Abstraction in Non-native Speech Perception. *Journal of Phonetics*, 46, 147–160.
- POLKA, L. (1995). Linguistic Influences in Adult Perception of Non-native Vowel Contrasts. *The Journal of the Acoustical Society of America*, 97(2), 1286–1296.
- SAITO, K., & POETEREN, K. VAN (2012). Pronunciation-specific Adjustment Strategies for Intelligibility in L2 Teacher Talk: Results and Implications of a Questionnaire Study. *Language Awareness*, 21(4), 369–385.
- SEERY, A.M., VOGEL-FARLEY, V., TAGER-FLUSBERG, H., & NELSON, C.A. (2013). Atypical Lateralization of ERP Response to Native and Non-native Speech in Infants at Risk for Autism Spectrum Disorder. *Developmental Cognitive Neuroscience*, 5, 10–24.
- THIESSEN, E.D., HILL, E.A., & SAFFRAN, J.R. (2005). Infant-directed Speech Facilitates Word Segmentation. *Infancy*, 7(1), 53–71. DOI: 10.1207/s15327078in0701\_5.
- TRAINOR, L.J., & DESJARDINS, R.N. (2002). Pitch Characteristics of Infant-directed Speech Affect Infants' Ability to Discriminate Vowels. *Psychonomic Bulletin & Review*, 9, 335. <https://doi.org/10.3758/BF03196290>.

- UTHER, M., KNOLL, M.A., & BURNHAM, D. (2007). Do You Speak E-NG-LI-SH? A Comparison of Foreigner- and Infant-directed Speech. *Speech Communication*, 49(1), 2-7.
- ZANGL, R., & MILLS, D.L. (2007). Increased Brain Activity to Infant-directed Speech in 6- and 13-month-old Infants. *Infancy*, 11, 31-62.