Proceedings of Meetings on Acoustics

Volume 19, 2013

<http://acousticalsociety.org/> **ICA 2013 Montreal****Montreal, Canada****2 - 7 June 2013****Speech Communication****Session 2aSC: Linking Perception and Production (Poster Session)**

2aSC11. Acoustic characteristics of Danish infant directed speech

Ocke-Schwen Bohn****Corresponding author's address: English, Aarhus University, Aarhus, DK-8000, none, Denmark, engosb@hum.au.dk**

Danish presents several challenges for language learners, such as a very densely packed upper portion of the acoustic vowel space, and a sibilant contrast that is acoustically less distinct than in e.g. English. The present study examined whether Danish caregivers enhance Danish contrasts when speaking to their 18 month old children (infant directed speech - IDS) as opposed to an adult (adult directed speech - ADS). Caregivers were recorded talking about toy animals in conversations with their child and with an adult interlocutor. The toy names were designed to elicit Danish contrasts differing in voice onset time and in place of articulation for sibilants, and vowels which are close neighbors in the crowded Danish vowel space. The dependent variables for the comparison of IDS to ADS were: VOT differences for homorganic stop consonants, the frequency at the amplitude peak for the sibilants, the Euclidean F1/F2 differences between vowels, F0 of the stressed (first) syllable in the toy name, as well as the duration of the stressed syllable, the vowels, and the fricatives. Results of the acoustic differences between ADS and IDS were compared to the results of parents' reports on the children's receptive vocabulary knowledge.

Published by the Acoustical Society of America through the American Institute of Physics

1. INTRODUCTION

It is well known that speakers may enhance phonetic contrasts when conversing with interlocutors who they assume to be linguistically not fully competent. This has been shown for a number of segment types and especially for vowels, which are hyperarticulated both in infant directed speech (IDS; Kuhl et al. 1997; Burnham, Kitamura & Vollmer-Conna 2002) and foreigner directed speech (FDS; Uther, Knoll & Burnham 2007). In both IDS and FDS, the corner vowels [i, a, u] are produced with more extreme articulations than in adult directed speech (ADS), resulting in larger acoustic vowel spaces and enhanced contrasts in IDS and FDS.

The present study is the first to examine Danish IDS. Specifically, this study addresses three questions:

1. Do Danish caregivers enhance Danish vowel and consonant contrasts when speaking to their child as opposed to an adult?
2. Does Danish IDS share global characteristics with IDS observed for other languages such as slower speech rate and higher pitch?
3. Is there a relationship between children's receptive vocabulary size and IDS characteristics?

1.1 Enhancement of Phonetic Contrasts

Danish presents several challenges for language learners and thus, opportunities for phonetic enhancement. As shown in Figure 1, Danish has an unusually large vowel inventory, and most Danish vowels are densely packed in the upper portion of the vowel space. Phonetic enhancement in IDS could result in more extreme values for the Danish corner vowels [i u, ɑ:] and larger acoustic distances between close neighbors in the vowel space. This study examined whether the Danish corner vowels [i u, ɑ:] are produced with more extreme acoustic values in IDS than ADS (i.e., lower F1 for [i, u] and higher F1 for [ɑ:], and higher F2 for [i] and lower F2 for [u].), and whether the acoustic difference between the close neighbors [e:ɛ:] and [o:ɔ:] would be enhanced (i.e., produced with a larger difference in the F1/F2 space) in IDS as compared to ADS.

FIGURE 1: Location of Danish vowels in the F1/F2 space. Data from Steinlen 2002 (mean values from 10 male speakers, 3 repetitions, /hVt/ context).

With respect to consonants, this study examined differences between IDS and ADS for two syllable-initial Danish contrasts: short-lag [t] vs. long-lag [t^h], and the sibilant contrast between [s] and [ç]. Guldager & Bohn (2013) report mean voice onset time (VOT) values for alveolar stop contrast (ADS produced in monosyllables in sentence context) of 21 ms for [t], and of 88 ms for [t^h]. One way in which this contrast could be enhanced would be through a reduction of the fairly long VOT for the short lag [t]. The Danish sibilant contrast between alveolar [s] and alveolo-palatal [ç] is characterized by mean spectral peaks of [s] and [ç] which are not as far apart as they are for the English and German sibilants [s] and [ʃ]: The Danish speakers in the present study produced [s] and [ç] in ADS with mean spectral peaks at 6549 Hz for [s] and 4537 Hz for [ç].¹

¹ Compare to Jongman, Wayland & Wong (2000), who report mean spectral peak locations for English [s,z] at 6839 Hz and for English [ʃ,ʒ] at 3820 Hz.

Since Danish has no [ʃ], the Danish [s] - [ç] contrast could be enhanced by lowering the spectral peak for [ç] (i.e., making it more [ʃ]-like by lengthening the front cavity).

1.2 Global Characteristics of IDS

In addition to acoustic measures related to specific Danish vowel and consonant contrast, the present study also examined whether previously reported global characteristics of IDS in other languages such as slower speech rate and higher pitch (Fernald 1992) would also be observed for Danish IDS. Slower speech rate, as reflected in longer syllable and vowel duration, may boost intelligibility (Sommers & Bracraft 2006), whereas higher pitch is related to positive affect (Burnham et al. 2002) and caregivers' attempts to direct infants' attention (Fernald 1985).

1.3 Relationship of IDS Characteristics to Receptive Vocabulary Size

Finally, this study attempts to throw light on a previously reported finding regarding Danish-learning children's early vocabulary development. Bleses et al. 2008a found that vocabulary comprehension scores for Danish-learning children (as elicited through the MacArthur-Bates Communicative Development Inventory – CDI, see Bleses et al. 2008b) were the lowest across 17 comparable CDI studies from age 1;0 onwards. For a subset of participants in the present study, vocabulary comprehension scores were obtained by parental report using a Danish adaptation of the CDI. These scores were compared to acoustic differences between IDS and ADS to examine whether a relationship existed between phonetic enhancement in IDS and size of the receptive vocabulary.

2. GENERAL METHODS

51 parent-child dyads, who had participated in a study on memory development (Kingo & Krøjgaard 2013) volunteered to participate in the current study. The parents, who were all native Danish speakers, reported that they spoke only Danish with their child.

The parent, the child, and an assistant (female, 24 years old) were seated in a sound booth (dimensions: 2m x 2m x 2m). The parent was informed that they were part of a study on language acquisition, and that their speech would be recorded. To ensure the use of the same words in IDS and ADS, the assistant gave the parent four soft toy animals (two at a time) onto which name tags had been fastened with made-up names which elicited the use of Danish segments of interest (see below).

For IDS, the parents were asked to talk to their child as they normally would when playing with a new toy, and to use the name of the stuffed animal rather than pronouns. The assistant presented each toy animal to the parent, saying the name of the animal out loud, and ensuring that the name tag was visible to the parent. When the child began to lose interest in the first two animals, the remaining two toys were introduced. Once the child began to lose interest in the last two animals, or became interested in something else, the assistant elicited ADS by asking the parent to describe the colors and names of the animals to the assistant rather than to the child. Throughout the IDS part of the recording session the assistant participated as little as possible, and interfered only encouragingly to keep the IDS conversation going.

The IDS and ADS conversations, which together lasted ca. 15 minutes, were recorded using a Zoom H2 digital stereo recorder and later downsampled to mono, 16-bit amplitude, 22050 Hz sampling rate, and segmented using Adobe Audition 3.0 (Adobe Systems 2007). All acoustic measurements were conducted in Praat (version 5.2.26, Boersma & Weenink 2011) using standard procedures. The number of target word tokens elicited in IDS and in ADS varied considerably. On average, ca. 15 tokens of each target word were produced by each parent in IDS conversations, of which the first 10 tokens were included in the analyses. Each parent produced ca. 2-3 tokens of each target word in ADS, all of which were included in the analyses.

2.1 Methods: Experiment 1

The focus of Experiment 1 was potential IDS-ADS differences with regard to a) the Danish [s] - [ç] contrast, b) the location and separation of Danish [i, e, u, a:] in the F1/F2 space. 33 parents were recorded in conversations with their child (IDS) and the assistant (ADS) using the four made-up toy names "Sjesser", "Sjisser", "Sjub", and "Sark". In Danish orthography, the initial letters <Sj> and <S> unambiguously represent [s] and [ç], and <i, e, u, a> unambiguously represent [i, e, u, a:] in these words.

The recordings from 7 (of 33) participants could not be used for further analyses.² In the 26 recordings which were analyzed, the child addressees (8 m, 18 f) had a mean age of 19 months. Acoustic analyses were conducted of the duration of

² Reasons for exclusion: Session had to be terminated because of child fussiness or because the speech signal from the parent could not be separated from child vocalizations.

the syllable (first syllable only in “Sjesser”, “Sjisser”), of the vowel and of the sibilant, of F0 at the target vowel midpoint, of the spectral peak location for the sibilant, and of F1 and F2 at the target vowel midpoint.

2.2 Methods: Experiment 2

The focus of Experiment 2 were potential IDS-ADS differences with regard to a) the Danish voicing contrast between syllable initial [t] and [t^h] in terms of VOT, b) the location and separation of Danish [o:-ɔ:] and [e:-ɛ:] in the F1/F2 space. 18 parents were recorded in conversations with their child (IDS) and the assistant (ADS) using the four made-up toy names “Doler”, “Dæler”, “Kåter” and “Teber”. In Danish orthography, the initial letters <D> and <T> unambiguously represent [t] and [t^h], and <o, æ, å, e> unambiguously represent [o:-, ɛ:-, ɔ:, e:] in these words.

The recordings from 5 (of 18) participants could not be used for further analyses.² In the 13 recordings which were analyzed, the child addressees (4 m, 9 f) had a mean age of 20.2 months. Acoustic analyses were conducted of the duration of the first syllable, of the target vowel, of VOT of the initial consonant, and of F0 and of F1 and F2 at the target vowel midpoint.

3. RESULTS: ACOUSTIC DIFFERENCES BETWEEN IDS AND ADS

This section first reports the global results on IDS-ADS differences from both Experiment 1 and Experiment 2 with respect to the duration of the (first) syllable of the target word and of the (first) vowel, and of F0 at the target vowel midpoint. Next, the results for the vowels will be presented and finally, the results for the consonants. Results were tested for significance using paired samples T-tests or, if the normality test (Shapiro-Wilk) failed, Wilcoxon Signed Rank tests.

3.1 Overall IDS-ADS Differences for Syllable and Vowel Duration, and for F0 in Target Syllable

Reports from previous studies (Fernald 1992) led to the expectation that the caregivers in the present study would use a slower speech rate in IDS than in ADS, and that the pitch of the target syllable would be higher in IDS than in ADS.

The mean duration of the first syllable in the target words was 315.7 ms (SD=55.1) in IDS, and 340.0 ms (SD=76.6) in ADS. A Wilcoxon Signed Rank test revealed that this difference was significant, $Z = 3.112$, $p = .002$. The mean duration of the first vowel in the target words was numerically shorter in IDS (122.6 ms, SD=21.3) than in ADS (127.3, SD=25.1), however, this difference was not statistically significant, $t(38) = -1.303$, $p = .200$. This suggests that, contrary to expectation, the parents did not speak more slowly when conversing with their child than when conversing with an adult.

The mean F0 in all target vowels was numerically lower in IDS (253.4 Hz, SD=39.0) than in ADS (259.5 Hz, SD=53.7). A Wilcoxon Signed Rank test revealed that this difference was not statistically significant, $Z = 1.326$, $p = .187$. Contrary to expectation, F0 in the target syllables was not higher in IDS than ADS.

3.2 IDS-ADS Differences for Vowels

To examine whether parents in the present study used more extreme vowel articulations when talking to their child than when talking to an adult, the triangle areas in the F1/F2 space defined by the three extreme vowels [i, u, ɑ:] in both speech styles were compared. The triangle areas were calculated from the side lengths of the triangle defined by the Euclidean distance in the F1/F2 space between [i] and [u], [i] and [ɑ:], and [u] and [ɑ:]. A paired-samples T-test revealed that IDS and ADS did not differ significantly in the size of the vowel triangle area, $t(20) = -0.635$, $p = .533$, suggesting that overall, parents did not use hyperarticulated vowels when talking to their children. Additional tests revealed that even though Euclidean distances in the F1/F2 space were larger in IDS than ADS for [i] to [ɑ:] and for [i] to [u], and smaller in IDS than ADS for [u] to [ɑ:], these distances were not statistically significant, $p > .2$.

Additional analyses were conducted to examine whether contrasts between closely neighboring vowels would be enhanced in IDS by producing these vowels with larger Euclidean distances in the F1/F2 space. For [i-e], this distance was nonsignificantly larger in IDS (mean=323.7) than ADS (mean=300.0), $t(20) = 0.561$, $p = .581$. Surprisingly, the Euclidean distance between [e:] and [ɛ:] was almost twice as large in ADS (208.0) than in IDS (107.1), which was statistically significant, $t(10) = -2.989$, $p = .014$. The Euclidean distance between [o:] and [ɔ:] was larger in ADS (202.8) than IDS (123.4), but a Wilcoxon Signed Rank test suggested that this difference was not significant, $Z = 1.156$, $p = .278$.

3.3 IDS-ADS Differences for Consonants [s], [ç] and [t], [t^h]

Sibilant productions in the two speech styles were analyzed for durational and spectral peak differences. A Wilcoxon Signed Rank test on the duration of sibilants produced in IDS (mean: 149.2 ms, SD=19.2) and in ADS (mean: 165.9 ms, SD=41.2) narrowly missed significance, $Z = 1.791$, $p = .075$. A Wilcoxon Signed Rank test examining the difference

between the duration of [s] in IDS (mean: 139.4 ms, SD=29.6) and in ADS (mean: 166.1 ms, SD=78.1) revealed no significance difference, $Z = 1.008$, $p = .322$). Surprisingly, the mean duration of the alveolo-palatal sibilant [ç] was significantly longer in ADS (172.6 ms) than in IDS (151.5 ms), $t(25) = -2.908$, $p = .008$.

To examine whether the parents would enhance the [s-ç] contrast, two analyses were carried out. The distance (in Hz) between the mean peak frequency locations of the sibilants was numerically larger in IDS (2328.5 Hz) than in ADS (2107.9 Hz), but a Wilcoxon Signed Rank test suggested that this difference was nonsignificant, $Z = -1.307$, $p = .202$. The peak frequency locations for [ç] did not differ significantly in IDS (4564.4 Hz) and ADS (4547.0 Hz), $t(25) = 0.209$, $p = .836$.

The comparison of the mean VOT difference between [t] and [t^h] revealed that this difference was significantly larger in ADS (89.1 ms) than in IDS (67.1 ms), $t(11) = -3.039$, $p = .011$. Follow-up tests found no significant differences between the VOTs for [t^h] in IDS and ADS, nor for the VOTs for [t] in the two speech styles, $p > .2$.

3.4 Conclusion

Table 1a and 1b summarize the comparisons of the acoustic characteristics of IDS and ADS. Surprisingly, relative phonetic enhancement was observed only in ADS. None of the acoustic parameters examined had a larger acoustic difference (in terms of duration or in terms of spectral properties) in IDS than in ADS. On the contrary, ADS was the clearer speech style with longer syllable duration, a larger duration for the sibilant [ç], a larger VOT difference for the [t- t^h] contrast, and a larger spectral difference between the neighboring vowels [e:] and [ɛ:].

TABLE 1a. Comparison of acoustic characteristics of IDS and ADS. A “+” sign in the row for IDS or ADS indicates phonetic enhancement of the acoustic parameter listed in the column head for either IDS or ADS; an “=” sign indicates that the acoustic difference between the two speech styles was not significant.

	duration		F0	vowel	Euclidean distance in F1/F2 space					
	syllable	vowel		triangle area	[i] - [a:]	[u] - [a:]	[i] - [u]	[i] - [e]	[e:] - [ɛ:]	[o:] - [ɔ:]
IDS		=	=	=	=	=	=	=		=
ADS	+								+	

TABLE 1b. Table 1a continued.

	duration		peak frequency		VOT			
	mean [s], [ç]	[s]	[ç]	difference [s] - [ç]	[ç]	difference [t] - [t ^h]	[t]	[t ^h]
IDS	=	=		=	=		=	=
ADS			+			+		

4. RESULTS: RELATION OF IDS TO RECEPTIVE VOCABULARY SIZE

For a subset of the participants in Experiment 1 (12 of 26) and Experiment 2 (12 of 13), parental reports on the vocabulary size and content were obtained using Danish version of MacArthur-Bates Communicative Development Inventory (CDI, Bleses et al. 2008b). Given that Danish-learning children’s vocabulary comprehension scores are low in comparison to children from other language backgrounds (Bleses et al. 2008a), it could be interesting to examine whether a relationship existed between children’s receptive vocabulary size and the degree of phonetic enhancement which their parents used in IDS. To address this issue, Pearson Product Moment correlations were analyzed comparing each child’s CDI score (receptive vocabulary) to the degree of phonetic enhancement in the parent’s IDS of the acoustic variables listed in Table 1. Phonetic enhancement is operationalized here as the signed difference between IDS and ADS for these variables.

Most of the acoustic variables were not related to the children’s receptive vocabulary score. Of the 18 variables examined, 14 had correlation coefficients with the CDI scores ranging from $r = .0111$ to $r = -.304$, $p > .34$. Three of these correlations are based on 24 data points each (syllable duration, vowel duration, and F0) and were clearly nonsignificant ($p < .62$). All remaining correlations are based on 10-12 data points each and should therefore be taken with a grain of salt, but it seems unlikely that additional data would have yielded significant results for correlations that, with the small number of data points in the present study, were far from reaching significance.

Two correlations revealed nonsignificant trends: Vowel triangle size x CDI score with $r = .491$, $p = .150$; and the Euclidean distance between [i] and [e] with $r = .480$, $p = .135$. These correlations are mentioned here because of the small number of data points (between 10-12) on which these analyses are based. It is not unlikely that data from additional child-parent dyads could strengthen the trend indicating a positive correlation between the size of the receptive vocabulary size and the two measures of vowel hyperarticulation. However, a marginally significant negative correlation between the [u] - [ɑ:] Euclidean distance and the size of the receptive vocabulary ($r = -.512$, $p = .089$) casts doubt on this possibility. This result seems to suggest (counterintuitively) that a smaller acoustic distance between the parent's articulation of the corner vowels [u] and [ɑ:] is related to a larger receptive vocabulary of the child. Still more puzzling is the only strong and significant correlation between the CDI score and acoustic enhancement in IDS: This score is clearly related to the acoustic distance between the corner vowels [i] and [u], $r = .797$, $p = .0032$.

5. DISCUSSION

The present study is the first to examine Danish Infant Directed Speech. Danish parents' conversation with their 19-20 month-old children (IDS) were compared to conversations of the same parents with an adult (ADS), and acoustic analyses were conducted to examine whether parents enhanced vowel and consonant contrasts of Danish when speaking to their child, and whether their speech rate would be slower, and their fundamental frequency higher, in IDS than in ADS. This study also examined whether a relationship exists between the children's reported size of the receptive vocabulary and the degree to which the parent's speech differences in IDS as opposed to ADS.

Surprisingly, this study revealed that the parents did not enhance phonetic contrasts or speak more slowly and use higher pitch when talking to their child than when talking to an adult. Further studies are needed to determine whether the lack of phonetic enhancement in Danish IDS is specific to Danish or to the age group (19-20 months) of children studied here. Still, this result casts doubt on the assumed universality of IDS characteristics (e.g., Fernald et al. 1989, Kuhl et al. 1997).

Even though most of the analyses of the relation between acoustics characteristics of parents' IDS and children's reported receptive vocabulary size were based on a small number data points, it can tentatively be concluded that global properties of Danish IDS such as speech rate and F0 are not related to the child's vocabulary development at 19-20 months of age, and that this is also true of phonetic characteristics of the parents' consonant productions as examined in this study. For some of the aspects of the parents' vowel productions, however, such a relationship was suggested by the data. Surprisingly, a marginally significant negative correlation was found for the Euclidean distance in the F1/F2 space between [u] and [ɑ:] and the CDI score, and a significant positive correlation was found for the Euclidean distance between [i] and [u] and the CDI score. It remains an open question for the time being why the acoustic distance between these corner vowel pairs (and not also [i] and [ɑ:]) should be related to size of the children's receptive vocabulary, and why the correlation should be negative for the [u] - [ɑ:] distance, and positive for the [i] - [u] distance.

ACKNOWLEDGMENTS

Research supported by the Danish National Research Foundation. Thanks to Peter Krøjgaard and Osman S. Kingo for sharing CDI data, to Stella B. Guldager for recruiting participants, running experiments, and analyzing data.

REFERENCES

- Blases, D. et al. 2008a. Early vocabulary development in Danish and other languages: A CDI-based comparison. *Journal of Child Language* 35, 619-650.
- Blases, D. et al. 2008b. The Danish Communicative Developmental Inventories: Validity and main developmental trends. *Journal of Child Language* 35, 651-669.
- Boersma, P. & Weenink, D. 2011. Praat: Doing phonetics by computer. Software downloaded from www.praat.org. (version 5.2.26)
- Burnham, D. K., Kitamura, C. & Vollmer-Conna, U. 2002. What's new, Pussycat? On talking to babies and animals. *Science* 296, 1435.
- Fernald, A. 1985. Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development* 8, 181-195.
- Fernald, A. 1992. Human maternal vocalizations to infants as biologically relevant signals: An evolutionary perspective. In: Barkow, J. H., Cosmides, L. & Tooby, J. (eds.), *The adapted mind*. New York: Oxford University Press, 391-428.
- Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B. & Fukui, I. 1989. A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language* 16, 477-501
- Guldager, S. B. & Bohn, O.-S. 2013. A study of voice onset time in Danish/English and English/Danish bilinguals' initial stop consonants. Poster prepared for the New Sounds 2013 Conference, Montreal, 17.6.-19.6.2013.

- Jongman, A., Wayland, R. & Wong, S. 2000. Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America* 108, 1252-1263.
- Kingo, O. S. & Krøjgaard, P. 2013. 18-month-olds' memory for short movies of simple stories. Ms. submitted for publication.
- Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U. & Lacerda, F. 1997. Cross-language analysis of phonetic units in language addressed to infants. *Science* 277, 684-686.
- Sommers, M. S. & Barcroft, J. 2006. Stimulus variability and the phonetic relevance hypothesis: Effects of variability in speaking style, fundamental frequency, and speaking rate on spoken word identification. *Journal of the Acoustical Society of America* 119, 2406-2416.
- Steinlen, A. K. 2002. The influence of consonants on native and non-native vowel production. Tübingen: G. Narr
- Uther, M., Knoll, M. A. & Burnham, D. 2007. Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication* 49, 2-7.