Hypoarticulation in infant directed speech

Kjellrun T. Englund

Department of Psychology, Norwegian University of Science and Technology

Contact:

Norwegian University of Science and Technology,

Department of Psychology

N-7491 Trondheim

Norway

Ph: +47 73 59 05 69

Mob: +47 92 44 20 89

E-mail: kjellrun.englund@ntnu.no

Abstract

A much-established finding in research on infant-directed speech (IDS) is that vowels are hyperarticulated compared to adult-directed speech (ADS). Studies showing this investigate point vowels, leaving us with a rather weak foundation for concluding whether IDS vowels are hyperarticulated within a particular language. The aim of this study was to investigate a large sample of vowels in IDS and to elicit speech in a natural situation for mother and infant. Acoustical and statistical analyses for /æ:, æ, ø:, ɵ, o:, ɔ, y:, y, ʉ:, ʉ, e:, ɛ/ show a selective increase in formant frequencies for some vowel qualities. In addition, vowels had higher fundamental frequency and were generally longer in IDS, but the difference between long and short vowels were comparable between IDS and ADS. With an additional front articulation and less lip protrusion in IDS compared to ADS, it is argued that IDS is hypoarticulated.

Keywords

Infant-directed speech, hypoarticulation, vowels, fronting

Introduction

One of the great puzzles of language acquisition is how infants learn phonetic contrasts at such an incredible speed. The phonetic learning that takes place during the first months is based on an infant's surrounding speech stimuli (Vallabha, McClelland, Pons, Werker, & Amano, 2007; Werker et al., 2007). It is widely believed that the speech infants receive has characteristics that facilitate learning, and we know that the speech register we use when interacting with an infant (infant-directed speech (IDS)) is different than the one we use when interacting with an adult (adult-directed speech (ADS)). Among other characteristics, the phonetic aspects of segments are different in IDS compared to ADS (for a review, see Cristia, 2013). For vowels, the vowel space is larger in IDS than ADS, indicating extreme articulation (Burnham, Kitamura, & Vollmer-Conna, 2002; Lam & Kitamura, 2008; Uther, Knoll, & Burnham, 2007). However, not all studies reveal the same pattern (Cristia & Seidl, 2013).

Some have shown discrepant findings with a smaller vowel space in IDS, and a shift for some vowel qualities (Benders, 2013; Englund & Behne, 2005). Research on vowels in IDS is mostly restricted to point vowels, perhaps providing us with a biased understanding of the facilitating input infants are thought to receive. In addition, most studies of IDS adopt a methodological approach where IDS is recorded once or only a few times (Benders, 2013; Green, Nip, Wilson, Mefferd, & Yunusova, 2010; Kuhl et al., 1997). Recording situations differ in the different studies, as do ages of the infants. This makes direct comparisons difficult and there is a need for studies that expand analyses to denser data sets, where larger parts of phonological inventories are studied and where numerous recordings are made of the same mother while ensuring a natural interactive setting to capture the true nature of the input.

This will clarify and broaden our view on the language environment that infants typically are surrounded with and learn from.

Hyperarticulation and hypoarticulation in IDS

In the Hyper-Hypo theory, IDS is viewed as an adaptation to a receiver who cannot predict the message very well (Lindblom, 1990). Under optimal listening conditions, and when predictability of the message is high, speech is relaxed with more assimilation. This is termed hypospeech. When, on the other hand, predictability is low and/or listening conditions are less than optimal, articulation becomes forceful with longer segments that are more audible, reducing ambiguity for the listener. This is called hyperspeech. A small infant does not have much linguistic experience and as a result, predictability will therefore almost always be low. When speaking to an infant we will consequently use hyperspeech, manifested by IDS.

The Hyper-Hypo theory relates to one of the predominant theories on phonological development, the Native Language Magnet (NLM) theory (Kuhl, 1993; Kuhl et al., 2008), which says that some prototypical vowel exemplars function as magnets for the perception of other exemplars. It is assumed that IDS vowels represent prototypical exemplars of vowel categories. Showing evidence of this is a study where American, Russian and Swedish mothers' IDS to their two- to five-month-old infants was analysed (Kuhl et al., 1997). In the languages studied, the vowels /a, i, u/ had generally more extreme F1 and F2 in IDS than in ADS, implying more extreme articulations. From these results it was suggested that extreme articulation makes IDS perceptually salient to the infant and aids language learning (Kuhl et al., 1997).

Hyperarticulated vowels are believed to aid language learning and indeed in Liu et al. (2003), a positive correlation between the size of mothers' vowel spaces in IDS and their 6- to 12-month-old infants' ability to discriminate vowels was demonstrated. An additional study has revealed that 21-month-old children learn words better from IDS than ADS (Ma, Golinkoff, Houston, & Hirsh-Pasek, 2011). Others have found strong evidence for the generality of vowel hyperarticulation as an instructive device for teaching language to learners (Uther et al., 2007). Hyperarticulation is modified by the degree of linguistic competence expected from the language learner (Xu, Burnham, Kitamura, & Vollmer-Conna, 2013), but see Burnham, Wieland, Kondaurova, McAuley, Bergeson and Dilley (2015), who show that vowel space characteristics are consistent across the first two years of an infant's life. Despite the extensive findings of hyperarticulation of vowels in IDS (Burnham et al., 2002; Lam & Kitamura, 2008; Liu, Kuhl, & Tsao, 2003; Xu et al., 2013), not all research points in the same direction.

There are studies displaying patterns of results more compatible with the hypoarticulation of IDS (Benders, 2013; Cristia & Seidl, 2013; Dodane & Al-Tamimi, 2007; Englund & Behne, 2005, 2006b; McMurray, Kovack-Lesh, Goodwin, & McEchron, 2013). Benders (2013) did a study of Dutch IDS using a paradigm in which mothers played freely with their 11- and 15-month-old infants using a set of selected toys to elicit words containing segments from the same phonetic surroundings. Findings included a small vowel space in IDS compared to ADS. In addition, mothers raised their F2 and F3 in corner vowels in IDS compared to ADS. The author points to these as acoustic markers of positive affect, rendering the idea of hyperarticulation as beneficial to phonological learning as secondary. A different finding was evident in Cristia and Seidl (2013), who conducted a study of American English IDS where the mothers' task was to describe objects to their children and to adults. In the study, mothers were asked to talk about categories to their infants and were provided objects/pictures of category exemplars to show to their babies. The study displayed different results for point vs. more central vowels. While vowel spaces for point vowels were expanded in IDS compared to ADS, [i-I] and [eI-ɛ] were not categorically separated

but had more overlap in IDS and not less overlap as would be expected within the view that mothers are trying to categorically separate speech segments. Therefore, although point vowels were produced with more peripheral acoustic characteristics in IDS than in ADS, hyperarticulation was not evident for phonemic differences other than place of articulation. The authors conclude that hyperarticulation is not a necessary feature of IDS. In another study of American English IDS by McMurray and co-workers, parents were recorded while reading a story to an infant and to an adult (McMurray et al., 2013). Findings revealed that while point vowels for the most part show a stretched vowel space, central vowels are not enhanced in IDS. Authors point to the large overlap between vowels and consequently question whether IDS enhances vowel category learning. They opt for a revision of the assumption that the content of IDS promotes language acquisition. In line with this are results from Dodane and Al-Tamimi (2007) who studied English, French and Japanese child-directed speech. They did not find a stretched vowel space, but rather a shift in the vowel triangle on the high-low dimension. Central vowels were more open, with higher F1 in IDS than in ADS. More specific are the findings for Norwegian by Englund and Behne (2006a), where hypoarticulation of point vowels was demonstrated with a reduced vowel space in IDS.

Due to the possibility that only some phonetic contrasts are enhanced in IDS, some researchers have pointed to the need for studying a broad range of contrasts in the same language (Martin et al., 2015; McMurray et al., 2013). This is also important in order to uncover if there are adaptations going on in IDS that can only be seen from a larger pattern of results, and not just from results based on just a few vowels. The approach of studying a broad set of vowels within the same language has the advantage of observing patterns of results that may have been hidden in previous studies.

In addition, the importance of a naturally occurring recording situation is that it will provide information on speech interaction that occurs with little or no instruction (Martin et al., 2015). However, Martin et al. (2015) used contrasts from the RIKEN corpus, which consists of IDS elicited when mothers were instructed to view picture books or engage in play. While being everyday activities, they were still induced by instruction, rendering less ecologically-valid recording situations. Situations that are initiated by the participants themselves in the comfort of their own homes will display the language input that an infant normally encounters. Together, this calls for studies of a broad set of contrasts, recording mothers with their small infants where they may feel most comfortable, in their own homes. This study is therefore an important contribution to enhance our understanding of early language acquisition.

Predictions

The knowledge we have to date of vowels in an infant's ambient language is much restricted to point vowels, not providing a full picture of the input from which an infant learns language. In addition, data collection is often restricted to a few vowels and low density in data as well as instructed elicitation of IDS. Higher density would increase precision of results and be a better foundation for conclusions. This invites studies that analyse fuller vowel inventories where mothers are recorded in natural situations to capture the true nature of the input.

The present study was designed to do this by looking at the large vowel inventory of Norwegian IDS. While three point vowel qualities have been studied before, six vowel qualities are unexplored and will be analysed in the current study. This will add to our knowledge of vowels in Norwegian IDS by covering the full vowel inventory. Most of the work on IDS is on vowel quality, but in Norwegian, vowel duration is a contrastive feature, corresponding to the term vowel quantity (Kristoffersen, 2000). With short and long vowels for each quality, Table 1 gives an overview of the 12 vowels included, exemplified by Norwegian minimal pairs.

Insert Table 1 here

From Table 1 it is clear that there are only two unrounded vowels /æ:, æ/ and /e:, ɛ/. The rest are rounded. Long and short vowels are represented within the same brackets; for example, when /y:, y/ is referred to throughout the paper, it is the vowel quality that is referred to, including both long /y:/ and short /y/. The studies showing hyperarticulation in IDS clearly outnumber those showing hypoarticulation (Cristia, 2013). If hyperarticulation is a general feature of IDS, then this should be evident also in the current study. As the vowels we study here are not point vowels, vowel space calculations are futile. Enhancement of contrastive features in vowels would ensure that one vowel should not stand the risk of becoming too similar to another vowel. Accordingly, hyperarticulation should render less overlap between vowel qualities in IDS compared to ADS. Although the relationship between formant frequencies and articulator movement is not a direct one, in general, when F1 decreases, the tongue has moved to a higher position and when F2 increases, the tongue has moved to a more front position. This means that increased F1 corresponds to a more open articulation, and increased F2 corresponds to a more front articulation. F3 corresponds to lip protrusion, which for rounded vowels would make them more distinct from unrounded vowels. The anterior-posterior dimension in the vocal tract has been closely tied to lip protrusion. The relationship between lip protrusion and F3 is inverse, with lower F3 with increasing protrusion (Kent & Read, 1992; Stevens, 1998). For the different vowel qualities, hyperarticulation therefore would imply:

- For /æː, æ/: More open (higher F1) in IDS than ADS.
- For /øː, ɵ/: Further back and more protruded lips (lower F2, lower F3) in IDS than ADS.
- For /oː, ɔ/: Further back, more closed and more lip protrusion (lower F2, lower F1 and lower F3) in IDS than ADS.
- For /yː, y/: Further back, more open and more lip protrusion (lower F2, higher F1 and lower F3) in IDS than ADS.
- For /ʉː, ʉ/: More closed and more lip protrusion (lower F1 and lower F3) in IDS than ADS.
- For /eː, ɛ/: More closed (lower F1) in IDS than ADS.

In addition, as Cristia points out that of 30 studies, 25 showed generally longer vowel duration or a reduced speech rate (Cristia, 2013), the longer duration of vowels in IDS is also expected here. Enhancement of contrastive phonetic features in IDS will additionally mean an interaction between speech type and quantity, where the difference between long and short vowels should be greater in IDS compared to ADS.

A generally higher F0 is an equally prevalent finding, with 33 out of 36 studies showing this in a metastudy (Cristia & Seidl, 2013), and a recent study confirming this for different language cultures (Broesch & Bryant, 2015); it is also expected here.

Method

The data for the current study come from a large corpus of natural speech[1], and details about data collection have previously been described (Englund & Behne, 2005). In the current research, care was taken to make recording settings as unobtrusive as possible to enable the spontaneous interaction that takes place in day-to-day activity and at the same time ensure the elicitation of IDS. This was accomplished by using a recording setting for IDS with direct face-to-face interaction between a mother and infant. The experimenter was present and interacted with the mother to elicit ADS, but was not present during IDS recordings. The mothers initiated the recordings themselves so that the situation came as close to everyday activity as possible.

Participants

Participants were enrolled from maternity groups at local health care centres and recordings started after their infants were born. Six native Norwegian-speaking mothers with a mean age of 27 years (range = 26 to 28 years) participated in the study. Their infants ranged from almost 4 to 24 weeks old. All mothers reported levels of education at bachelor level or higher. Mothers and infants were generally healthy throughout the study. Upon introduction, mothers signed an informed consent and after completing the study, they were briefed about the purpose of the study and received sound files with their own IDS recordings.

Procedure and equipment

Recordings were made over a 6-month period. A headset microphone (SHURE, model WH20) with a frequency response from 50 to 15000 Hz was connected to a Sony Digital Audio Tape recorder Walkman TCD–D8 for recording both ADS and IDS. For ADS, two

[1] Data collection for the present project is approved by the Committee for Medical and Health Research Ethics and registered in the Norwegian Social Scientific data register.

headsets ran through a Behringer Eurorack MX602 mixer. Each recording session included both IDS and ADS recordings, and for a typical IDS recording the mother and infant were alone in the room, while ADS was recorded in a conversation between the experimenter and the mother, usually in the living room. Recording time varied and ranged from approximately 10 to 45 minutes. A typical IDS recording was 15 minutes, and a typical ADS recording was 30 minutes. Each mother was instructed to change the infant's nappy, interacting with her child as she would normally do in an everyday situation. Other than that, no instructions were given. ADS recordings were natural conversations about anything the mother initiated as a topic. The development of her infant was a recurring topic, as were general news items from papers etc. At the beginning of an ADS recording, the mother was asked if she remembered any of the words she used while making the IDS recording. In this way, some words occurred in both IDS and ADS. Typically, instructions were given only at the first and second recordings, as it seemed artificial to repeat them.

Acoustic analyses

The current research was conducted in order to explore a wider range of vowel qualities, therefore including the vowel qualities /æ:, æ, ø:, ɵ, o:, ɔ, y:, y, ʉ:, ʉ, e:, ɛ/. A further aim was to explore a representative sample of phonetic contexts for each vowel. Consequently, all occurrences of target vowels in content words were from words in focal position in a sentence. It has been found that compared to content words, vowel durations in function words are longer in IDS than ADS (Bernstein Ratner, 1985). In addition, hyperarticulation may be different for words in focal position (Martin, Utsugi, & Mazuka, 2014). Therefore, the same percentage of content and function words were sampled from ADS as from IDS, from the different vowel qualities and quantities. With a corpus of natural speech, the vowels used for measurements occurred in a variety of phonetic contexts. Cases where vowels preceded or followed a liquid, glide or nasal, from which it can be difficult to distinguish the vowel, were kept to a minimum.

PRAAT (Boersma & Weenink, 2009) was used to conduct acoustic analyses. Visual inspection initiated determination of the beginning and end of a vowel, and was supplemented by auditory judgement. Duration was measured in milliseconds. For formants, each measurement was based on the mean for all frames whose centres lie within the selected time span. Means of the first, second and third formant frequencies were calculated in Hertz for the total selected frame. If vowels were not visibly evident in the spectrum, in cases of background noise, where the speaker had a creaky voice, or when there was a heavy puff of air during articulation, the vowel was rejected from further analyses. From the hours of recordings available, 3028 segments were analysed. Selection depended on only one speaker being audible, as well as no noise on the recording. As everyday activity includes the use of objects, running water etc., this considerably reduced the number of words that were feasible for analyses. Sentences were transcribed for further analysis. Care was taken to include vowels where the start and end points could be determined from periods with considerable amplitude.

The study ran over a period of six months. Some hesitated to start the study with their newly born infants, resulting in a varying start of the first recordings. In addition, mothers 2 and 6 did not complete the last recording. Consequently, there was unevenness in data density for different time points. When in addition, a previous study from the same corpus of IDS showed no changes in vowel spectra in IDS over the first six months (Englund and Behne, 2006) data was collapsed for analysis. Results were analysed by the IBM SPSS (version 21.0) statistical package. Duration was measured in milliseconds but, as duration is perceived logarithmically, a transformation was applied to duration values. The statistical software uses a natural logarithm, returning base-e logarithm of the duration values. Analyses were run for the recalculated variable (Kondaurova, Bergeson, & Dilley, 2012). The Mel scale takes the nonlinearity of frequency perception into consideration (Stevens, Volkmann, & Newman, 1937). Fundamental frequency and formant frequencies were recalculated by using the formula (m = 2595 log 10 (1 + f / 700)) from O'Shaughnessy (2000).

Results

Independent variables were speech type with two levels (ADS or IDS), vowel quality with six levels /æ:, æ; ø:, ɵ; o:, ɔ; y:, y; ʉ:, ʉ, e:, ɛ/, and vowel quantity with two levels (long and short). The Mahalanobis procedure for detecting outliers was employed, and after removing extreme values at +/- 3 SD at either side of the mean, ADS had 1529 tokens and IDS 1236 tokens. N for the different mothers included in the analyses was: for mother 1: 648, mother 2: 387, mother 3: 495, mother 4: 378, mother 5: 401 and mother 6: 456. This led to the following distribution over speech qualities: For /æ:, æ/, n = 435; for /ø:, ɵ/, n = 229; for /o:, ɔ/, n = 505; for /y:, y/, n = 290; for /ʉ:, ʉ/, n = 494; for /e:, ɛ/, n = 812; with 1046 long vowels and 1719 short vowels. Means and standard deviations for variables F1, F2 and F3 in Mels are presented in Table 2.

Insert Table 2 here

Table 2 shows means and standard deviations for F1 – F3 in Mels for vowel qualities in ADS and IDS. From Table 2 it is apparent that IDS has generally higher standard deviations than ADS. This was followed up by computing standard deviations into new variables for all vowel qualities in ADS and IDS, and running repeated measures ANOVAs for logduration, F0, F1, F2 and F3. Analyses revealed that in all dependent variables except logduration, standard deviations were higher for IDS than for ADS. Means, standard error as well as results from repeated measures analyses are presented in Table 3.

Insert Table 3 here

Table 3 shows that the standard deviations were significantly higher in IDS than ADS for F1, F2 and F0. From the analyses two interactions between speech type and vowel quality emerged. For F2: F (5, 25) = 25.87, p = .000, and for F3: F (5, 25) = 10.91, p = .000. Paired samples t-tests revealed that for F3, the only two vowel qualities where IDS had higher standard deviations than ADS were /y:, y/ (t (5) = - 4.19, p = .009), and /e:, ɛ/ (t (5) = - 7.23, p = .001). For F2, standard deviation was higher in IDS than ADS for /y:, y/, /ʉ:, ʉ/ and /e:, ɛ/ (t (5) = -3.49, p = .017, t (5) = - 4.53, p = .006, and t (5) = - 4.52, p = .006, respectively). For F2 there was one instance where ADS had significantly higher standard deviation than IDS (t (5) = 7.82, p = .001).

Insert Figure 1 and 2 here

Figure 1 shows F1 – F2 representation of all vowel qualities in ADS, and Figure 2 shows the same for IDS. From these figures, what seems to be a shift in vowel space appears with a vowel distribution that is higher on F2 in IDS than ADS. In addition, a large variation for the different vowel qualities is evident.

Linear mixed models analyses were carried out for dependent variables ‘F0’, ‘log duration’, ‘F1’, ‘F2’ and ‘F3’. In the model, fixed effects were ‘speech type’ with two levels, ‘vowel quality' with six levels and 'vowel quantity' with two levels. Cluster variable were 'subject'. Since in a mixed model interactions are test values against the highest of the values for the variable in question, for vowel quality this meant that each vowel quality was tested separately with a similar model where interactions appeared. As natural speech is used in the current study, the degree of inherent variability is necessarily high and a 5% level of significance was used. Table 4 shows main effects (tests of fixed effects) of the dependent variables from mixed models analyses.

Insert Table 4 here

The last two predictions of a generally longer duration and higher F0 in IDS than ADS were supported (Table 4). Table 2 shows that the significantly higher log duration in IDS compared to ADS seen from the means also shows a different result for some vowel qualities. The significant interaction between speech type and segment showed that for /æ:, æ/ and /o:, ɔ/, IDS vowels were longer than ADS vowels [F (1, 3025) = 23.82, p = .000 and F = (1, 3025) = 12.34, p = .000, respectively]. However, for /ø:, ɵ/, /y:, y/, /ʉ:, ʉ/ and /e:, ɛ/, it was the other way round: ADS vowels were longer than IDS vowels [F (1, 3031) = 102.2, p = .000; F (1, 3052) = 61.60, p = .000; F (1, 3050) = 67.89, p = .000; and F (1, 3030) = 49.65, p = .000, respectively]. Log duration was reliably higher for long vowels (Table 2) than short vowels and this was stable across speech types [F (1, 3032) = 1.52, p = n.s.]. F0 was significantly higher in IDS than ADS and further investigation into the interaction between speech type and vowel quality (Table 4), showed that only for /y:, y/ was F0 higher in IDS than ADS [F (1, 3019) = 4.61, p = .032].

For /æ:, æ/, the prediction of a higher F1in IDS (Table 2) was not supported. However, further analyses based on the significant interaction between speech type and vowel quality showed that for /æ:, æ/, the minimally higher F1 in ADS compared to IDS is significant [F (1, 3054) = 77.08, p = .000]. F1 was also significantly different between IDS and ADS for /ø:, ɵ/, /o:, ɔ/, /y:, y/ and /e:, ɛ/ [F (1, 3059) = 27.00, p = .000; F (1, 3051) = 16.09, p = .000; F (1, 3057) = 19.17, p = .000; and F (1, 3059) = 33.59, p = .000, respectively]. It was higher in IDS for /o:, ɔ/ but for the other three vowel qualities it was lower in IDS. The only case where F1 did not differ between the speech types was /ʉ:, ʉ/ [F (1, 3058) = 2.56, n.s.].

The second prediction of a lower F2 in IDS than ADS for /ø:, ɵ/ (Table 2) was not supported. On the contrary, a higher F2 in IDS was observed (Table 2) [F (1, 3058) = 24.86, p = .000]. For the other vowel qualities, the significant interaction between speech type and vowel quality revealed F2 to be no higher in IDS for /æ:, æ/, /y:, y/ or /ʉ:, ʉ/ [F (1, 3058) = 2.80, n.s.; F (1, 3056) = 0.60, n.s.; F (1, 3058) = 2.77, n.s., respectively], but for /o:, ɔ/ and /e:, ɛ/ [F (1, 3058) = 6.74, p = .009; F (1, 3056) = 49.65, p = .000] (see Table 2).

For /o:, ɔ/, the expected lower F1 and F2 in IDS was thereby not supported. Neither was the expected lower F2 and higher F1 in IDS than ADS for /y:, y/. For /ʉ:, ʉ/, the expectation was a lower F1 in IDS than ADS, but this was not evident from the analyses. However, the lower F1 for IDS compared to ADS in /e:, ɛ/ was supported.

A generally higher F3 appeared in IDS compared to ADS (Table 2), and further analyses of the significant interaction between speech type and vowel quality (Table 4) showed there to be no speech type difference in F3 for /e:, ɛ/ [F (1, 3057) = 0.39, p = .528]. However, a higher F3 in IDS appeared for /æ:, æ/, /ø:, ɵ/, /o:, ɔ/, /y:, y/ and /ʉ:, ʉ/ [F (1, 3059) = 21.44, p = .000; F (1, 3059) = 22.19, p = .000; F (1, 3059) = 38.05, p = .000; F (1, 3057) = 10.41, p = .001; and F (1, 3058) = 16.24, p = .000, respectively] (see Table 2).

Discussion

Summary of results

This study aimed at broadening our knowledge of IDS to young infants by studying the large vowel inventory of Norwegian IDS and recorded mothers at home in everyday situations while their infants were very young. The following is an overview of predictions, with comments as to whether each was supported.

- More open /æ:, æ/ in IDS. Not supported, the effect was the opposite.
- More back and protruded /ø:, ɵ/ in IDS. Not supported, the effect was the opposite.
- More back, closed and protruded /o:, ɔ/ in IDS. Not supported, the effect was the opposite.
- More back, open and protruded /y:, y/ in IDS. Not supported, no effect (the opposite for protrusion).
- More closed and protruded /ʉ:, ʉ/ in IDS. Not supported, no effect (the opposite for protrusion).
- More closed /e:, ɛ/ in IDS. Supported.
- Longer duration IDS. Supported.
- Higher F0 in IDS. Supported.

The following effects seem evident from the current results: Vowel qualities /ø:, ɵ/, /o:, ɔ/ and /e:, ɛ/ are more fronted; /o:, ɔ/ is more open; /ø:, ɵ/, /y:, y/and /e:, ɛ/ are less open; and, in addition, vowel qualities /æ:, æ/, /ø:, ɵ/, /o:, ɔ/, /y:, y/ and /ʉ:, ʉ/ are less protruded in IDS compared to ADS. The results correspond to those of Benders (2013), with raised F2 and F3, and parts of McMurray et al. (2013) and Cristia and Seidl (2013), with a larger overlap between vowel contrasts. They also accord with Dodane and Al-Tamimi (2007) and Englund and Behne (Englund & Behne, 2005), who observed a shift of some vowels in the front-back dimension. Based on previous and current findings, the questions become: what is the mother doing in IDS and why is she doing so?

Acoustic – articulatory relations

Some ask whether adaptations in IDS are secondary to higher fundamental frequency or reduced speaking rate (McMurray et al., 2013). We know that infants preference for IDS (Cooper, Abraham, Berman, & Staska, 1997) is related mainly to F0 (Segal & Newman, 2015). Higher F0 in IDS could be the result of an attempt to mimic infant production (Cristia, 2013). When we know that infants prefer to listen to utterances produced by other infants (Masapollo, Polka, & Ménard, 2015), one would get an infant's attention by doing this. Due to smaller vocal tracts, this would lead to higher F0, but also more open and more front vowels (Ménard, Schwartz, Boë, Kandel, & Vallée, 2002). However, both emotionality and mimicking speech would affect all vowel qualities in a similar way. The selective increase in formants for some qualities in this study does not support either explanation.

Speech rate has been used to explain the usual longer vowel duration in IDS. Infants easily attend to more slow IDS which is high in affect (Panneton, Kitamura, Mattock, & Burnham, 2006), and slow speech improves word recognition (Song, Demuth, & Morgan, 2010). This is not surprising, seeing that in slow speech there is a decreased probability of target undershoot (Gay, Ushijima, Hirose, & Cooper, 1974). In addition, a slow speaking rate would lead to a more open jaw, and more time to reach articulatory extremes (Vanson & Pols, 1990). Although certainly part of what is going on in IDS, also here with longer vowel log duration in IDS, neither higher pitch nor slower speech can alone account for the selective adjustment in formants observed for some vowel qualities in the present data.

Varying articulatory parameters does not always lead to the same acoustic result (Stevens, 1998) and one should be cautious not to interpret the connections between acoustic measures and articulatory movement too directly. Nevertheless, some connections are seen between the position and movement of the articulators and the acoustic outcome. Generally, the first formant frequency is linked to jaw opening. Jaw opening was generally comparable between speech types, but mothers articulated /o:, ɔ/ with a more open jaw in IDS. But the fact that the opposite was true for /ø:, ɵ/, /y:, y/ and /e:, ɛ/ leads to a conclusion that vowels are neither articulated with a more open nor a more closed jaw in IDS.

The second formant is sensitive to placement of the tongue body (Kent & Read, 1992; Stevens, 1998), and the third formant is sensitive to placement of the tip of the tongue (Sundberg, 1977). F2 and F3 can be seen in connection, and if the tongue body is more front, it is probable that the tip of the tongue will be equally front. Therefore, where F2 is increased, it is likely that F3 is correspondingly increased and that lips are more protruded. The current results showed three vowel qualities to be more front and five vowel qualities to be less protruded in IDS than ADS. The lack of more opening is surprising, seeing that infants have an 'F1 bias', i.e., the finding that young infants are better at discriminating vowel contrasts conveyed by F1 than contrasts that are associated with corresponding F2 changes (Lacerda & Sundberg, 2001). This, together with selective fronting and less rounding in vowels, paints a picture where IDS represents less-specified vowels.

Hypoarticulation as perceptual challenge

As observed in Table 3, and confirmed through analyses of standard deviations, although more so for some vowel qualities, variation is generally higher in IDS than ADS. Together with lack of specification of vowel quantity, lack of jaw opening, and selective fronting as well as less-rounded vowels the current results do not support hyperarticulation but rather point more in the direction of hypoarticulation. Hypoarticulation in IDS does not coincide well with the traditional idea that this speech type enhances speech category learning. Instead, it justifies the idea that IDS may be a perceptual challenge.

Large variation is also mentioned as a central finding in McMurray et al. (2013) and there is also a possibility that high variability in IDS is what makes it challenging, and that the hypoarticulation observed is a masked effect of variability where large variability leads to overlapping phonetic categories. Statistical learning models are based on an estimate of both the mean and variability in values that characterise phonemes. If variability is high, this may outweigh the benefit of expanding the vowel space to establish prototypes. In this way, highly variable IDS clearly entails a perceptual challenge to an infant who is faced with the task of learning phonetic categories. This goes against findings that computer models learn speech contrasts better from hyperarticulated (Boer & Kuhl, 2003). However, it should be mentioned that not all agree with what de Boer and Kuhl concluded (Kirchhoff & Schimmel, 2005)and when, further, some have found the opposite, namely that in fact ADS outperforms IDS as a foundation for learning of some contrasts (McMurray et al., 2013), we have to consider the possibility that although IDS may be a perceptual challenge, it may still not be detrimental to phonetic learning. Instead, such a challenge may be beneficial to a speech-learning infant. But how can it be beneficial? Work on categorisation of meaning offers a useful analogue. At the heart of stimuli processing in infants, is the idea that low levels of variability and complexity might lead to habituation, which in turn causes low attention and counteracts learning. In this way, variation is not necessarily harmful to categorisation. A study of visual categorisation has shown that variability is central in defining category membership (Mather & Plunkett, 2011). Note that habituation would be happening only if a stimulus was presented repeatedly, and the referred study used 10-month-olds and with a very different purpose. Still, a perceptual challenge hypothesis could be set forth for phonetic learning with the equivalent idea, presupposing variation to be necessary for attention and learning in phonetic category development. A new study supports this hypothesis, by using a mathematical teaching model. Although not all learners can profit from variability, large variability may lead to better learning by directing the learners inferences away from segments that are not good exemplars, and towards segments that are (Eaves, Feldman, Griffiths, & Shafto, 2016). Additional research comes from studies of second language acquisition in adults where it is shown that spectrally more variable materials from different talkers leads to learning of more robust categories (Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994; Sadakata & McQueen, 2013; Wong, 2014). Infants also learn words faster when presented with multiple speakers (Rost & McMurray, 2009), but also from one speaker with varying duration, overall pitch and pitch contour (Galle, Apfelbaum, & McMurray, 2015). Infants may search for invariant cues in the speech they encounter, whether auditory or visual, and invariants may become evident if variants outnumber them.

Visual perceptual aspects

Some articulatory gestures may have acoustic effects that are more or less easy to perceive and the content of IDS may or may not be intentional by the mother. Regardless, why would mothers speak with selectively more fronted vowels and less lip protrusion? Much of the speech that infants encounter is multimodal. We know that infants prefer to look at infant-directed faces (Kim & Johnson, 2014), and a study has shown that infants as young as two months old perceive the audio-visual aspects of sounds within syllables (Baier, 2007). If visual cues are important in language learning, they may also be fundamental to IDS and, although it would be expected as a general effect, the observed fronting in IDS could be motivated by enhancing visual speech cues to infants. One highly visible aspect of speech would be jaw opening, but this was not observed as an aspect of the current IDS. Another would be fronting, which could mean vowel articulatory movements in some way would be easier for an infant to see. When only half of the vowel qualities were fronted (/æ:, æ/, /o:, ɔ/, /e:, ε/) and at least two of the ones that were already central vowels, it is not easy to grasp in what way it would make the visual task easier for an infant. A third highly visible aspect in vowel production is lip protrusion. The effect of protruding the lips is the lengthening of the vocal tract and lowering of formant frequencies. However, this was not the case here. In fact, in the current data, it seems most vowels are less protruded. A related possibility is that IDS represents increased visual contrastiveness between categories. Although some have questioned it (Ter Schure, Junge, & Boersma, 2016), a study by Teinonen and coworkers shows that visual information aids phoneme learning in infancy (Teinonen, Aslin, Alku, & Csibra, 2008).

Smiling

There is one explanation which may cover the selective fronting and less lip rounding observed in the current data. Revealing emotional information, the mother could be smiling while producing IDS. If a speaker is smiling, the tongue body is more front and lips are less protruded (Tartter, 1980). In addition to making speech cues visible, it could increase infant production, as shown in a study where the quantity of speech-like syllabic infant vocalizations increases if the mother is smiling during face-to-face interaction (Hsu, Fogel, & Messinger, 2001). Although the study was done with adults, a study has also reported that listeners may attach more weight to visual input from a smiling rather than austere speaker (Traunmuller & Ohrstrom, 2007). When smiling, the mouth widens and the lips retract resulting in a shortened vocal tract with a resulting increase in all formants. However, it seems to have different consequences for rounded vs. unrounded vowels. A study has shown that for the vowel /u:/, a smile resulted in significantly higher F3, while for /a:/ and /i:/ it did not. So, lip protrusion decreased more from a smile if the vowel was inherently more protruded (Fagel, 2010). This coincides well with the current data, where there were no difference in F3 for /eː, ɛ/.

The smiling explanation could be further supported by the choice of recording situation. The current approach used a face-to-face interactive setting, which may have encouraged smiling. This is not a common setting to use in IDS research and could explain the different results in a study of Swedish, where a play situation was used to elicit IDS (Kuhl et al., 1997). Studying infants under 6 months challenges the adaptation of interactive situations, and with a 2-month-old it may be unnatural to play on the floor with toys, while with a 6-month-old it may be more natural. In this way, the recording situation should be adapted to the age and development of the infant. The recording situation selected for the current study ensured inclusion of infants from birth and to six months. As the mother and infant were alone in the current setting it is seen as highly natural and representative for natural occurring IDS, but it may also have affected the kind of adaptations the mother is using in that particular situation.

Methodological aspects

Accounts of IDS rarely discuss the possibility that the recording situation may have a profound effect on experimental results. In Green et al. (2010), the point was made that the lack of more articulatory exaggeration can be due to self-consciousness while being observed. Although some have shown no difference (Stern, Spieker, Barnett, & MacKain, 1983), it has also been shown that some mothers speak more slowly in a home setting than in a laboratory setting (Stevenson, Leavitt, Roach, Chapman, & Miller, 1986). If this is so, one can question whether mothers use more extreme IDS at home and this would mean that the slower speaking rate should be evident here, but the lack of longer vowels in IDS refutes this. With a lab-oriented approach with older infants, there might be less face-to-face interaction, and one may observe less focus on visual speech cues. Research is underway to test whether characteristics of Norwegian IDS changes in line with recording situations.

Interpreting the impact of what the mother does during IDS recordings depends on what she does during the ADS recording. Recent work has pointed to the possibility that the differences between speech registers is in part due to the nature of ADS recordings (Johnson, Lahey, Ernestus, & Cutler, 2013). Most studies of IDS use ADS speech where the adult is unfamiliar to the mother, typically an experimenter. Johnson et al. (2013) have shown that differences between ADS to a familiar adult and IDS are smaller than those between ADS to an unfamiliar adult and IDS, which may lead to a bias when interpreting the characteristics of IDS. In the current methodological approach, each mother spoke to the same experimenter at 12 points in time in the families' homes. So, although the bias may have been relevant for the first couple of recording sessions, the experimenter and mothers became increasingly friendly throughout the study. In consequence, the effect is likely evened out by later recording sessions. While this might explain if reduced register effects were found (e.g., no hyperarticulation), it does not predict significant differences when they do occur. This gives the current study an advantage with a valid ADS condition, containing speech that comes close to what mothers would use with other familiar/friendly adults.

Summary

The early language environment may present considerable complexity to infants who are about to learn phonetic categories. The present study was designed as a thorough approach studying an abundance of vowels collected in IDS and ADS in a natural interactive setting early in infant life. Data provides a striking picture of results showing vowels to be hypoarticulated and selectively open, fronted and less protruded in IDS. While hypoarticulation may complicate an infant's auditory language learning, it may also facilitate perception of visual aspects of speech and emotional aspects in communication. Results call for theoretical development in IDS research that acknowledges that within the emotional and attention-getting message of IDS lies a perceptual challenge for an infant.

References

Baier, R., Idsardi, W.J. and Lidz, J. (2007). *Two-month-olds are sensitive to lip rounding in dynamic and static speech events.* Paper presented at the International Conference of Audio-Visual Speech Processing, Casteel, Groenendael, Hilvarenbeek, The Netherlands.

Benders, T. (2013). Mommy is only happy! Dutch mothers' realisation of speech sounds in infant-directed speech expresses emotion, not didactic intent. *Infant Behavior & Development, 36*(4), 847-862. doi:10.1016/j.infbeh.2013.09.001

Bernstein Ratner, N. (1985). Dissociations between vowel durations and formant frequency characteristics. *Journal of Speech and Hearing Research, 28*, 255-264.

Boer, B. d., & Kuhl, P. K. (2003). Investigating the role of infant-directed speech with a computer model. *Acoustics Research Letters On-line, 4*(4), 129–134.

Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer (version 5.3.69) Downloaded at: http://www.praat.org/.

Broesch, T. L., & Bryant, G. A. (2015). Prosody in infant-directed speech is similar across Western and traditional cultures. *Journal of Cognition and Development, 16*(1), 31-43. doi:10.1080/15248372.2013.833923

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new pussycat? On talking to babies and animals. *Science, 296*.

Burnham, E. B., Wieland, E. A., Kondaurova, M. V., McAuley, J. D., Bergeson, T. R., & Dilley, L. C. (2015). Phonetic modification of vowel space in storybook speech to infants up to 2 years of age. *Journal of Speech, Language, and Hearing Research, 58*(2), 241-253. doi:10.1044/2015_jslhr-s-13-0205

Cooper, R. P., Abraham, J., Berman, S., & Staska, M. (1997). The development of infants' preference for motherese. *Infant Behavior & Development;Infant Behavior & Development, 20*(4), 477-488. doi:10.1016/s0163-6383(97)90037-0

Cristia, A. (2013). Input to language: The phonetics and perception of infant-directed speech. *Language and Linguistics Compass, 7*(3), 157-170.

Cristia, A., & Seidl, A. (2013). The hyperarticulation hypothesis of infant-directed speech. *Journal of Child Language, 41(4)*, 913-934.

Dodane, C., & Al-Tamimi, J. (2007). *An acoustic comparison of vowel systems in adult-directed speech and child-directed speech: Evidence from French, English and Japanese*. Paper presented at the 16th International Congress of Phonetic Sciences, Saarbrucken, Germany.

Eaves, B. S., Feldman, N. H., Griffiths, T. L., & Shafto, P. (2016). Infant-Directed Speech Is Consistent With Teaching. *Psychological Review, 123*(6), 758-771. doi:10.1037/rev0000031

Englund, K., & Behne, D. (2006a). Changes in Infant Directed Speech in the First Six Months. *Infant and Child Development, 15*(2), 139-160.

Englund, K. T., & Behne, D. M. (2005). Infant directed speech in natural interaction - Norwegian vowel quantity and quality. *Journal of Psycholinguistic Research, 34*, 259-280.

Englund, K. T., & Behne, D. M. (2006b). Changes in infant directed speech in the first six months. *Infant and Child Development, 15*(2), 139-160.

Fagel, S. (2010). Effects of Smiling on Articulation: Lips, Larynx and Acoustics. In A. Esposito, N. Campbell, C. Vogel, A. Hussain, & A. Nijholt (Eds.), *Development of Multimodal Interfaces: Active Listing and Synchrony* (Vol. 5967, pp. 294-303).

Galle, M. E., Apfelbaum, K. S., & McMurray, B. (2015). The Role of Single Talker Acoustic Variation in Early Word Learning. *Language Learning and Development, 11*(1), 66-79. doi:10.1080/15475441.2014.895249

Gay, T., Ushijima, T., Hirose, H., & Cooper, F. S. (1974). Effect of speaking rate on labial consonant-vowel articulation. *Journal of Phonetics, 2*, 47-63.

Green, J. R., Nip, I. S., Wilson, E. M., Mefferd, A. S., & Yunusova, Y. (2010). Lip movement exaggerations during infant-directed speech. *Journal of Speech, Language, and Hearing Research, 53*(6), 1529-1542.

Hsu, H. C., Fogel, A., & Messinger, D. S. (2001). Infant non-distress vocalization during mother-infant face-to-face interaction: Factors associated with quantitative and qualitative differences. *Infant Behavior & Development, 24*, 107-128.

Johnson, E. K., Lahey, M., Ernestus, M., & Cutler, A. (2013). A multimodal corpus of speech to infant and adult listeners. *Journal of the Acoustical Society of America, 134*(6), EL534-EL540. doi:10.1121/1.4828977

Kent, R., & Read, C. (1992). *The acoustic analysis of speech*. San Diego, California: Singular Publishing Group.

Kim, H. I., & Johnson, S. P. (2014). Detecting "infant-directedness" in face and voice. *Developmental Science, 17*(4), 621-627.

Kirchhoff, K., & Schimmel, S. (2005). Statistical properties of infant-directed versus adult-directed speech: Insights from speech recognition. *Journal of the Acoustical Society of America, 117*(4), 2238-2246. doi:10.1121/1.1869172

Kondaurova, M. V., Bergeson, T. R., & Dilley, L. C. (2012). Effects of deafness on acoustic characteristics of American English tense/lax vowels in maternal speech to infants. *Journal of the Acoustical Society of America, 132*(2), 1039-1049. doi:10.1121/1.4728169

Kristoffersen, G. (2000). *The phonology of Norwegian*. Oxford: Oxford University Press.

Kuhl, P. K. (1993). Innate predispositions and the effects of experience in speech perception: The Native Language Magnet theory. In p. nate, B. d. B.-B. the effects of experience in speech perception: The Native Language Magnet theory, S. d. Schonen, P. W. Jusczyk, P. McNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 259-274). New York, NY, US: Kluwer Academic/Plenum Publishers.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., . . . Lacerda, F. (1997). Crosslanguage analysis of phonetic units in language addressed to infants. *Science, 277*, 684-686.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of The Royal Society B, 363*, 979-1000.

Lacerda, F., & Sundberg, U. (2001). Biases in early language acquisition. In F. Lacerda, C. v. Hofsten, & M. Heimann (Eds.), *Emerging cognitive abilities in early infancy*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Lam, C., & Kitamura, C. (2008). *"Your baby can't hear you": How mothers talk to infants with simulated hearing loss*, Brisbane, QLD.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H & H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403-439): Kluwer.

Liu, H. M., Kuhl, P. K., & Tsao, F. M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science, 6*(3), F1-F10.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America, 96*(4), 2076-2087. doi:10.1121/1.410149

Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word learning in infant- and adult-directed speech. *Language Learning and Development;Language Learning and Development, 7*(3), 185-201. doi:10.1080/15475441.2011.579839

Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., & Cristia, A. (2015). Mothers speak less clearly to infants than to adults: A comprehensive test of the hyperarticulation hypothesis. *Psychological Science, 26*(3), 341-347. doi:10.1177/0956797614562453

Martin, A., Utsugi, A., & Mazuka, R. (2014). The multidimensional nature of hyperspeech: Evidence from Japanese vowel devoicing. *Cognition, 132*(2), 216-228. doi:10.1016/j.cognition.2014.04.003

Masapollo, M., Polka, L., & Ménard, L. (2015). When infants talk, infants listen: Pre-babbling infants prefer listening to speech with infant vocal properties. *Developmental Science*, No Pagination Specified. doi:10.1111/desc.12298

Mather, E., & Plunkett, K. (2011). Same items, different order: Effects of temporal variability on infant categorization. *Cognition, 119*(3), 438-447. doi:10.1016/j.cognition.2011.02.008

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*(2), 362-378. doi:10.1016/j.cognition.2013.07.015

Ménard, L., Schwartz, J.-L., Boë, L.-J., Kandel, S., & Vallée, N. (2002). Auditory normalization of French vowels synthesized by an articulatory model simulating growth from birth to adulthood. *Journal of the Acoustical Society of America, 111*(4), 1892-1905. doi:10.1121/1.1459467

O'Shaughnessy, D. (2000). *Speech Communication: human and machine*: Addison Wesley.

Panneton, R., Kitamura, C., Mattock, K., & Burnham, D. (2006). Slow Speech Enhances Younger But Not Older Infants' Perception of Vocal Emotion. *Research in Human Development, 3*(1), 7-19. doi:10.1207/s15427617rhd0301_2

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science, 12*(2), 339-349. doi:10.1111/j.1467-7687.2008.00786.x

Sadakata, M., & McQueen, J. M. (2013). High stimulus variability in nonnative speech learning supports formation of abstract categories: Evidence from Japanese geminates. *Journal of the Acoustical Society of America, 134*(2), 1324-1335. doi:10.1121/1.4812767

Segal, J., & Newman, R. S. (2015). Infant preferences for structural and prosodic properties of infant-directed speech in the second year of life. *Infancy, 20*(3), 339-351. doi:10.1111/infa.12077

Song, J. Y., Demuth, K., & Morgan, J. (2010). Effects of the acoustic properties of infant-directed speech on infant word recognition. *Journal of the Acoustical Society of America, 128*(1), 389-400. doi:10.1121/1.3419786

Stern, D. N., Spieker, R. K., Barnett, R. K., & MacKain, K. (1983). The prosody of maternal speech: infant age and context related changes. *Child Language, 10*(1), 1-15.

Stevens, K. N. (1998). *ACOUSTIC PHONETICS*: Cambridge: Massachusetts Instit Technology Press.

Stevens, S. S., Volkmann, J., & Newman, E. B. (1937). The mel scale equates the magnitude of perceived differences in pitch at different frequencies. *Journal of the Acoustical Society of America, 8*(185).
Stevenson, M. B., Leavitt, L. A., Roach, M. A., Chapman, R. S., & Miller, J. F. (1986). Mothers' speech to their 1-year-old infants in home and laboratory settings. *Journal of Psycholinguistic Research, 15*(5), 451-461. doi:10.1007/bf01067725
Sundberg, J. (1977). The Acoustics of the Singing Voice. *Scientific American*.
Tartter, V. C. (1980). Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception and Psychophysics, 27*, 24-27.
Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition, 108*(3), 850-855.
Ter Schure, S., Junge, C., & Boersma, P. (2016). Discriminating Non-native Vowels on the Basis of Multimodal, Auditory or Visual Information: Effects on Infants' Looking Patterns and Discrimination. *Frontiers in Psychology, 7*. doi:10.3389/fpsyg.2016.00525
Traunmuller, H., & Ohrstrom, N. (2007). Audiovisual perception of openness and lip rounding in front vowels. *Journal of Phonetics, 35*(2), 244-258. doi:10.1016/j.wocn.2006.03.002
Uther, M., Knoll, M., & Burnham, D. (2007). Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication, 49*(1), 2-7.
Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 104*(33), 13273-13278.
Vanson, S. E., & Pols, L. C. W. (1990). Formant frequencies of Dutch vowels in a text, read at normal and fast rate. *Journal of the Acoustical Society of America, 88*(4), 1683-1693. doi:10.1121/1.400243
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.
Wong, J. W. S. (2014). The Effects of High and Low Variability Phonetic Training on the Perception and Production of English Vowels /e/-/ae/ by Cantonese ESL Learners with High and Low L2 Proficiency Levels. In H. Li & P. Ching (Eds.), *15th Annual Conference of the International Speech Communication Association* (pp. 524-528).
Xu, N., Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2013). Vowel hyperarticulation in parrot-, dog- and infant- directed speech. *Anthrozoös, 26*(3), 373-380. doi:10.2752/175303713x13697429463592

Table 1

Word 1	Long vowel	Word 2	Short vowel
Være (be)	[æ:]	Verre (worse)	[æ]
Søt (cute)	[ø:]	Søtt (sweet)	[ɵ]
Våt (wet)	[o:]	Vått (wet, neuter)	[ɔ]
Syn (vision)	[y:]	Synd (shame)	[y]
Lun (snug)	[ʉ:]	Lund (grove)	[ʉ]
Sen (late)	[e:]	Send (send)	[ɛ]

Table 1. Overview over the vowels under study exemplified by minimal pairs. The first and fourth column represent the two words in the minimal pair. English meanings are added in parentheses. Column two and five represent the corresponding long and short vowels in these minimal pairs in IPA.

Table 2

Mean and (SD) for F0, F1, F2, F3 (mel), duration and logduration for long and short vowels in IDS and ADS									
		Dur long	Dur short	Ldur long	Ldur short	F0	F1	F2	F3
Infant-directed speech	**[æ:] [æ]**	169(129)	71(27)	4.9(.59)	4.1(.37)	367(95)	764(72)	1519(111)	1974(91)
	[ø:] [ɵ]	147(104)	65(38)	4.8(.57)	4.0(.50)	360(88)	622(95)	1605(122)	2003(80)
	[o:] [ɔ]	285(234)	68(31)	5.3(.75)	4.1(.42)	374(92)	671(119)	1357(145)	2006(84)
	[y:] [y]	97(51)	52(19)	4.4(.45)	3.9(.33)	348(75)	504(85)	1688(128)	2024(68)
	[ʉ:] [ʉ]	103(64)	61(27)	4.5(.49)	4.0(.40)	362(86)	543(102)	1610(126)	1992(78)
	[e:] [ɛ]	125(63)	62(27)	4.7(.47)	4.1(.40)	367(81)	626(116)	1615(132)	2019(79)
		Dur long	**Dur short**	**Ldur long**	**Ldur short**	**F0**	**F1**	**F2**	**F3**
Adult-directed speech	**[æ:] [æ]**	99(88)	64(26)	4.4(.48)	4.1(.33)	305(58)	769(86)	1423(90)	1966(72)
	[ø:] [ɵ]	126(64)	58(21)	4.7(.46)	3.9(.37)	308(63)	628(70)	1437(92)	1923(68)
	[o:] [ɔ]	208(194)	61(33)	4.9(.82)	3.9(.47)	300(52)	629(88)	1236(123)	1950(68)
	[y:] [y]	108(67)	50(18)	4.5(.51)	3.8(.33)	307(49)	519(64)	1637(106)	1951(60)
	[ʉ:] [ʉ]	138(133)	53(20)	4.7(.60)	3.9(.36)	324(45)	523(63)	1476(87)	1893(67)
	[e:] [ɛ]	124(68)	64(30)	4.7(.46)	4.0(.40)	300(58)	675(81)	1508(110)	1973(63)

Table 2. Means and standard deviations for Duration (msec), logduration, F0 - F3 (Mels) for vowel qualities in ADS and IDS.

Table 3

Standard deviations for logduration, F0, F1-F3 in ADS and IDS			
	ADS	**IDS**	**F-value (df), alphalevel**
logdur	.20	.22	F (1,5) = .889, p = .389
F1	43.8	52.8	F (1,5) = 21.86, p = .005
F2	57.8	68.7	F (1,5) = 7.02, p = .044
F3	29.7	33.7	F (1,5) = 2.01, p = .215
F0	21.5	31.5	F (1,5) = 17.88, p = .008

Table 3. Standard deviations for logduration, F0, F1 - F3 for both ADS and IDS, and results from a repeated measures analyses for these dependent variables with speech type and vowel quality as independent variables.

Table 4

F-values and p-levels for fixed effects and interactions					
	Logdur	**F0 (mel)**	**F1 (mel)**	**F2 (mel)**	**F3 (mel)**
Speech	F (1, 3033) = 15.3, p = .000	F (1, 3003) = 303.4, p = .000	F (1, 3039) = 0.7, p = .793	F (1, 3039) = 410.9, p = .000	F (1, 3039) = 263.8, p = .000
Quality	F (5, 3033) = 35.5, p = .000	F (5, 3003) = 0.8, p= .523	F (5, 3039) = 408.0, p = .000	F (5, 3039) = 482.2, p = .000	F (5, 3039) = 35.6, p = .000
Quantity	F (1, 3033) = 1063.4, p = .000	F (1, 3003) = 0.1, p = .741	F (1, 3039) = 40.2, p = .000	F (1, 3039) = 0.1, p = .768	F (1, 3039) = 2.7, p = .101
Speech X Quality	F (5, 3033) = 10.2, p = .000	F (5, 3003) = 3.2, p = .007	F (5, 3039) = 15.6, p = .000	F (5, 3039) = 7.1, p = .000	F (5, 3039) = 16.7, p = .000

Table 4. F-values and p-levels from mixed model analyses for main effects and interactions for vowel log duration, F0, F1, F2, F3.

Figure 1

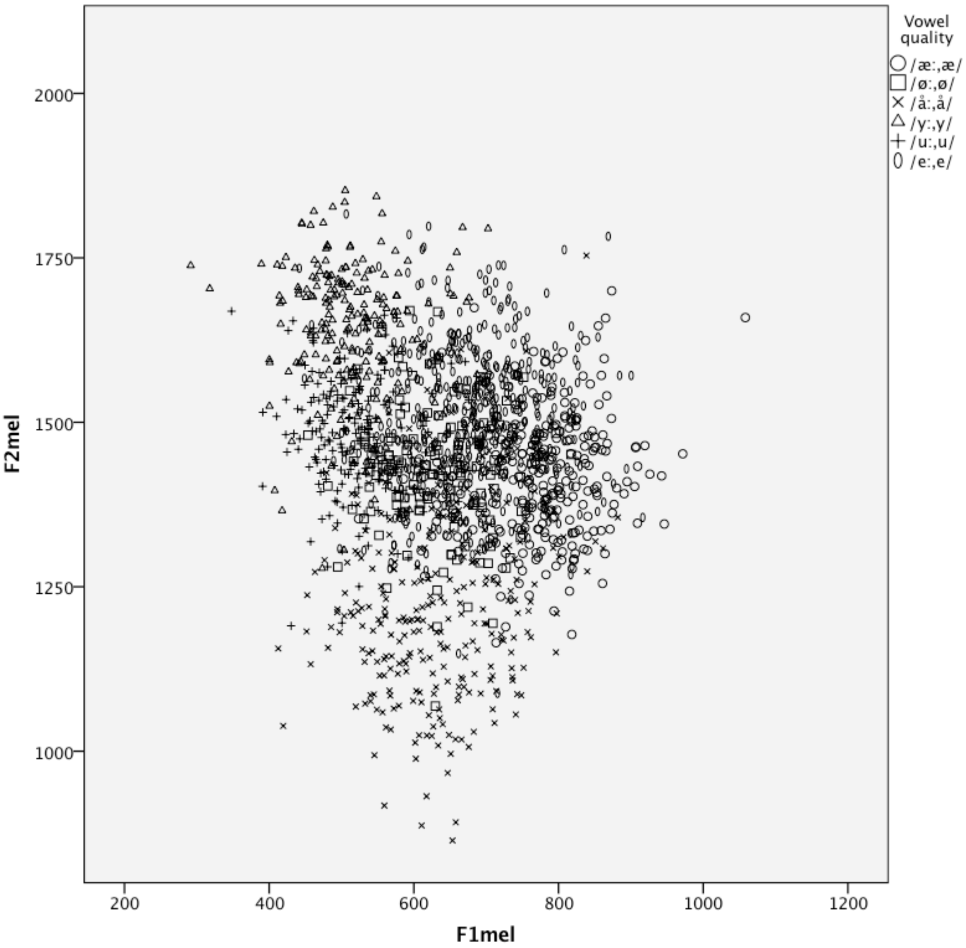

Figure 1. F1-F2 distribution in Mels for ADS. Each point represents one segment.

Figure 2

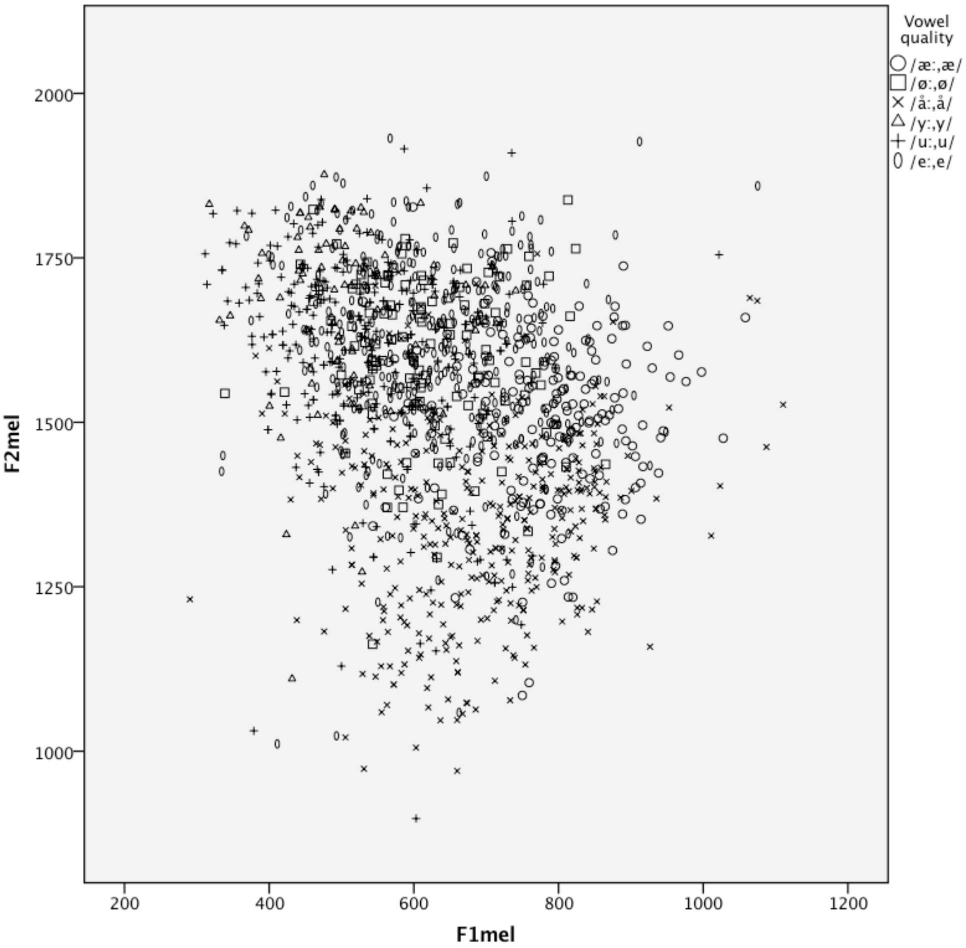

Figure 2. F1-F2 distribution in Mels for IDS. Each point represents one segment.