



# Acoustic and Affective Comparisons of Natural and Imaginary Infant-, Foreigner- and Adult-directed Speech

<sup>1</sup>Monja Knoll, <sup>1,2</sup>Lisa Scharrer

<sup>1</sup>Ecological Acoustic Research Group, Dept. Psychology, University of Portsmouth, UK

<sup>2</sup>Psychologisches Institut, Ruprecht-Karls-Universität Heidelberg, Germany

monja.knoll@port.ac.uk

## Abstract

This study evaluated the use of imagined interactions in speech research, by comparing speech addressed to imaginary speech partners with natural speech addressed to genuine interaction partners. Samples of speech directed to an imaginary infant (IDS), foreigner (FDS) and adult (ADS) produced by ten female students were acoustically analysed and also rated on positive vocal affect. Our results for vocal affect are consistent with previous findings using natural interactions, with IDS rated higher in positive vocal affect than ADS/FDS. However, acoustic analyses of IDS revealed a much smaller vowel space than ADS/FDS, with no difference between those two conditions. Unlike the findings in the natural speech samples, our IDS mean pitch was not significantly higher than ADS/FDS. Since these results are contrary to those from interactions with genuine speech partners, speech obtained from imaginary interactions should be used with caution.

**Index terms:** IDS, imaginary speech, hyperarticulation

## 1. Introduction

Infant-directed speech (IDS) is acoustically and phonetically modified compared to adult-directed speech (ADS) [e.g. 1, 2]. The well recognised characteristics of IDS include exaggerated pitch contours, increased mean pitch, hyperarticulation and high emotional affect [e.g. 3, 4, 5]. Researchers now believe that acoustic modifications in IDS probably have a linguistic role in language acquisition, but that they might also have emotional-attentional functions [e.g. 1, 3]. A growing body of research has attempted to separate these functions by comparing IDS with other linguistic (foreign-directed speech: FDS) [6] or emotional affective groups (e.g. pets or partners) [3, 5]. A variety of different methodologies have been used in these investigations. For instance, IDS has been compared to ADS using students and imaginary scenarios in the laboratory [7, 8], and in natural interactions in the mother's home [5, 6].

A number of advantages and limitations are inherent in each of these methodologies. Laboratory conditions offer greater potential to control the environment with regards to noise levels, room acoustics and consistency of recordings. Additionally, using students imagining talking to relevant groups is more time efficient and convenient than using genuine interaction partners. In the case of IDS and FDS, it might be particularly difficult to find an infant or foreign confederate that is available over a prolonged period of time. The potential disadvantages of laboratory studies are that the resulting interactions might be contrived and unnatural. For instance, people's ability to imagine talking to certain listener

groups might depend both on their previous exposure to these groups, and their 'acting' ability. Secondly, it is not quite clear how comparable the findings of these studies are to genuine interactions, as relevant comparison studies are rare [e.g. 3, 9]. Conversely, studies using natural interactions in the home environment can be extremely time intensive, and the recording set-up has to be carefully arranged and taken into account in the evaluation of the acoustic analyses. The advantages of natural interactions in the home environment, however, are that the interactions are genuine rather than contrived.

Considering these differences, there is clearly a need for a comparison study of both natural and laboratory based IDS research. This is particularly so as two recent studies [6, 7] used different methodologies to investigate differences between IDS, FDS and a British adult control group (ADS). Biersack *et al.* [7] used an imaginary laboratory scenario with student speakers, whereas a previous study by the first author and others [Uther *et al.*; 6] used natural interactions with mother speakers. With regards to mean pitch, both studies reported similar findings in that IDS achieved higher pitch than FDS. However, Biersack *et al.*'s [7] study reported findings on speech rate and pitch range modification, whereas Uther *et al.*'s [6] study concentrated on examining hyperarticulation of vowels and emotional affect. As such it might be difficult to compare these two studies given their different methodologies. A study that attempted to compare spontaneous speech (mothers with their infants) with non-spontaneous speech (students with an imaginary interlocutor) was carried out by Schaeffler *et al.* [9]. However, despite describing their study as comparing spontaneous versus non-spontaneous interactions, Schaeffler *et al.* [9] still used standardised sentences in brochures for both types of interactions. Therefore, although their study represents an important first step in comparing non-spontaneous to spontaneous speech, it is still unclear to which extent genuine natural interactions are comparable to imaginary student interactions.

Here we intended to replicate Uther *et al.*'s [6] study in the laboratory with student speakers, and with the help of imaginary scenarios. In contrast to Schaeffler *et al.*'s [9] study, however, we did not provide our speakers with brochures and standardised sentences, but provided them with the same toys (sheep, shark & shoe) and a description of the 'imaginary' confederates used in Uther *et al.* [6]. Uther *et al.* [6] had found that both IDS and FDS were characterised by hyperarticulation of vowels compared to ADS. However, both adult conditions exhibited significantly lower mean pitch and vowel duration than IDS, with no significant differences between the adult conditions. Not surprisingly, the authors found that IDS was also characterised by the presence of high emotional affect (obtained via ratings of the mothers' low-pass

filtered speech) compared to the two adult conditions. However, they also found that FDS was perceived to contain less positive vocal affect (more negative vocal affect) than ADS. Given that our study differed from the original Uther *et al.* [6] study in only the use of imaginary speech partners, we would expect to find broadly similar results if imagined interactions are truly ecologically valid.

## 2. Method

The speech samples of Uther *et al.* [6] consisted of 10 southern English mothers (mean age 30.7) in interactions with their infants, a foreign (Chinese) and southern English confederate (both adult females). For each of the interactions the mothers were provided with three toys to keep consistency of conversation content, otherwise the interactions were natural and spontaneous. More information about the mothers, confederates and procedures can be found in Uther *et al.* [6]. However, apart from using recordings of imaginary speech produced by students instead of natural speech produced by mothers, all our procedures follow Uther *et al.* [6], and are detailed below.

### 2.1. Participants

#### 2.1.1. Speakers (students)

The speakers were 10 females with a mean age of 22.9 years (*sd* 8.84) recruited from the student population of the University of Portsmouth. As in the Uther *et al.* [6] study, all speakers were British citizens with comparable southern English accents, who had been living in the south-east of England for most of their lives (> 16 years).

#### 2.1.2. Listeners (raters)

For the ratings of vocal affect, we recruited 20 students via the participant pool of the Department of Psychology, University of Portsmouth. Participants consisted of 8 males and 12 females with a mean age of 27.25 (*sd* 10.37). None of the participants were hearing impaired, and all were British citizens. These participants were required to rate low-pass filtered speech samples of the student speakers for vocal affect. Inter-rater reliability was found to be high (reliability coefficient  $\alpha = .80$  for positive vocal affect; reliability coefficient  $\alpha = .78$  for negative vocal affect).

### 2.2. Acoustic analyses and filtering

Overall, 346 words were analysed. Acoustic analyses were carried out using Praat 4.5.16, [10] and centred on the target words shark, sheep and shoe containing the corner vowels /a/, /i/ and /u/. The resulting sound samples were analysed for mean fundamental frequency (F0), mean vowel duration, and formant 1 and 2 (F1/F2). F1/F2 values were used to plot vowel triangles for /a/, /i/ and /u/, from which vowel triangle area was calculated to detect 'hyperarticulation' of the vowels.

For the affective analysis, 20 seconds of each speech sample were low-pass filtered (frequency pass Hann band) with a cut off point of 1000 Hz, and smoothing at 100 Hz. We used this level of low-pass filtering to keep our study comparable to Uther *et al.* [6], who also used 1000 Hz. Otherwise, the speech samples received no further modification (see additional material for examples). Ideally, low-pass filtering should remove the intelligibility of speech without impairing its affective features, and the procedure has recently been used in IDS research [e.g.1, 5].

### 2.3. Procedure

For each of the procedures (speakers and listeners), informed consent was obtained before the experiment began, and debriefing was given at the end of the experiment.

#### 2.3.1. Speech samples

The speakers were required to imagine talking to an infant, a British adult and a foreign adult in separate interactions. For the infant interaction, they were instructed to imagine that they were talking to a close family member (e.g. niece, nephew or even own child). However, we did not provide speakers with an example or idea of what IDS would sound like. For both the British and foreign adult interaction, they were instructed to imagine talking to a female stranger in her early twenties. Additionally, in the foreign interactions they were instructed to imagine that the person had been living in the UK for less than two months, and that they might encounter some communication problems. To keep these interactions consistent with Uther *et al.* [6] and to elicit the same target words, we supplied the speakers with three toys (a shark, a sheep and a shoe). The speakers were encouraged to use these toys for the interactions (e.g. to play with them with the imaginary infant, or explain which toy they would buy for an infant in the adult interactions). Apart from the toy stimuli, the speakers were encouraged to construct and invent their own scenarios, which ideally should have resulted in free speech. The order of the interactions for each speaker was counterbalanced, and each interaction lasted for approximately two minutes.

#### 2.3.2. Affective ratings

Raters listened to 30 low-pass filtered speech samples (10 each of IDS, FDS and ADS). Using a five point Likert-scale (1 (not at all) to 5 (extremely)), they were required to rate the speech samples on both positive and negative vocal affect. The speech samples were randomised (using Microsoft Office Excel 2003) and counterbalanced. Each participant was supplied with two test trials of the filtered speech to familiarise themselves with the filtered sound.

## 3. Results

Our main aim was to investigate whether we could replicate the findings of Uther *et al.*'s [6] natural speech study with a data set based on imaginary interactions. It should be noted here that ADS was used as a type of baseline condition in the original study, whereas FDS was a linguistic comparison group to IDS. The data for the acoustic analyses were subjected to a repeated measures ANOVA, whereas the data for the affective scales (positive and negative vocal affect) was submitted to a two factor repeated ANOVA with speech type (IDS, ADS and FDS) and the ten different speakers as the within subjects factors.

### 3.1. Acoustic analyses

We found a significant difference for vowel space (hyperarticulation) between the three speech types ( $F(2, 18) = 4.036, p = .036, \eta = .31$ ). IDS vowel space was significantly smaller than both ADS ( $F(1, 9) = 6.103, p = .036$ ) and FDS ( $F(1, 9) = 9.192, p = .014$ ) vowel space (see Fig. 1). We found no significant difference between either adult condition. This is in contrast to Uther *et al.*'s [6] study, where no significant difference between IDS and FDS vowel space was found, but both conditions presented significantly greater vowel spaces than ADS (see Fig. 1 for comparison of vowel space of both studies). Similar to Uther *et al.* [6], we

compared the vowel duration between IDS, FDS and ADS, however in contrast to their study (IDS > ADS > FDS) we found no significant difference between the speech types and mean duration differences were minimal (< 10 ms).

The difference of mean pitch (fundamental frequency) between the speech types approached significance ( $F(2, 18) = 3.308, p = .06, \eta = .27$ ), so we calculated planned contrast between the three speech types. IDS achieved significantly higher mean pitch than ADS ( $F(1, 9) = 5.657, p = .041$ ), but there was no significant difference for mean pitch between IDS and FDS. The mean pitch difference between FDS and ADS only approached significance ( $F(1, 9) = 4.922, p = .054$ ) with FDS presenting higher mean pitch than ADS (see Fig. 2).

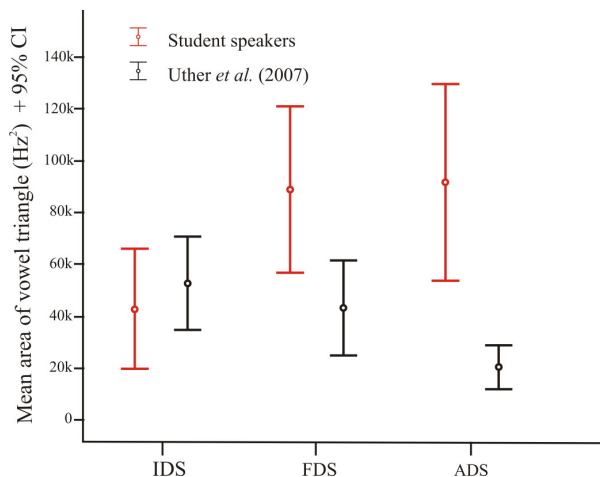


Figure 1: Representation of IDS, FDS and ADS mean vowel space (calculated from vowel triangle in F1/F2 space) for student speakers in comparison with speakers in Uther et al. [6]. Error bars represent 95% confidence interval.

In contrast to our findings, Uther et al. [6] had found that IDS exhibited significantly higher mean pitch than both adult conditions, with no significant difference between ADS and FDS (see Fig. 2). Overall, the mean difference between IDS and the two adult conditions in our study is lower than expected (< 30 Hz), and the mean pitch of IDS is considerably lower than in previous studies [e.g. 1, 3].

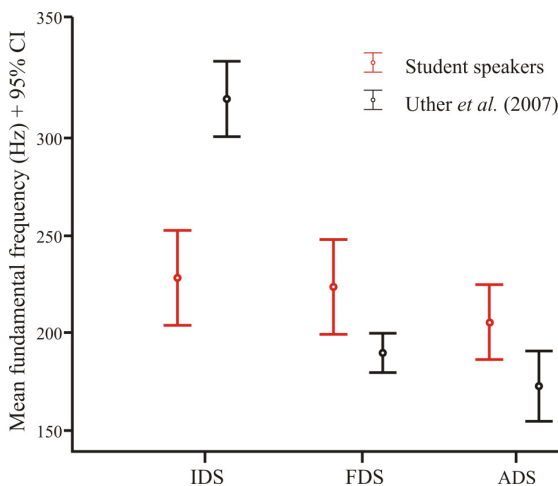


Figure 2: Mean pitch for IDS, FDS and ADS for student speakers in comparison with speakers in Uther et al. [6]. Error bars represent 95% confidence interval.

### 3.1.1. Affective analyses

We also investigated whether we could replicate Uther et al.'s [6] findings with regards to ratings of emotional affect (positive and negative vocal affect). We found a significant difference for ratings of positive vocal affect between the speech types ( $F(2, 38) = 17.158, p < .001, \eta = .475$ ) with IDS obtaining significantly higher ratings of positive vocal affect than both ADS ( $F(1, 19) = 13.723, p = .002$ ) and FDS ( $F(1, 19) = 28.952, p < .001$ ). The difference between ADS and FDS only approached significance ( $F(1, 19) = 11.045, p = 0.054$ ), with ADS achieving higher ratings of positive vocal affect than FDS (see Fig. 3). Not surprisingly, the results for negative vocal affect follow a similar trend, and a significant difference between the speech types was found ( $F(2, 38) = 16.120, p < .001, \eta = .459$ ). IDS achieved significantly lower ratings of negative vocal affect than both ADS ( $F(1, 19) = 6.028, p = .024$ ) and FDS ( $F(1, 19) = 24.143, p < .001$ ). Complementary to the positive vocal affect, FDS achieved lower ratings of negative vocal affect than ADS, but here the difference was significant ( $F(1, 19) = 16.712, p = .001$ ).

With regards to emotional affect, our findings are comparable to those of Uther et al. [6], who found that IDS achieved significantly higher ratings of positive vocal affect (lower ratings of negative vocal affect) than both adult conditions, with ADS achieving higher ratings of positive vocal affect than FDS (see Fig. 3).

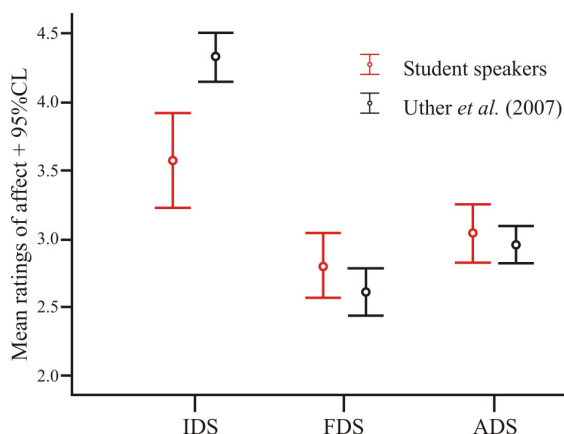


Figure 3: Representation of positive vocal affect for each of the speech types compared with Uther et al. [6]. Ratings of negative vocal affect are not presented as they follow the same trend. Error bars represent 95% confidence intervals.

## 4. Discussion

This study set out to evaluate the validity of IDS research in natural and laboratory conditions. Our results suggest that, while some of the acoustic and affective aspects of the different speech types can be reproduced in a laboratory context, others cannot. For instance we were unable to replicate the occurrence of hyperarticulation in both IDS and FDS, which was one of the main findings of Uther et al.'s [6] study. If we consider ADS as a normal baseline speech type, our findings are particularly interesting as the students in our study in effect *hypoarticulated* in IDS (i.e. they produced a smaller vowel space than in ADS or FDS). This finding is difficult to explain, but may relate to the speakers' limited experience with infants and lack of feedback from the 'imaginary' infant.

It is notable that the speakers in the imaginary ADS and FDS reported that their 'dialogue' was uttered in an explanatory fashion, which focused on the target words. In

contrast, in IDS the speakers tended to bracket the target words with typical ‘baby sounds’, and made less of an attempt to teach the words to their imaginary infant compared to mothers in the original study. This observation is also supported by the fact that our mean vowel space for ADS and FDS is much higher than in the natural interaction. This finding cannot be explained purely by individual or regional differences, as we chose speakers with the same regional accents as those used by Uther *et al.* [6]. The speakers may therefore have used clearer (hyperarticulated) speech in the two adult conditions as part of a more instructional ‘teacher’ manner of speaking.

Our findings with regards to mean pitch are also interesting. Although we found increased mean pitch in IDS compared to ADS, overall our IDS mean pitch is considerably lower than that of Uther *et al.* [6] and other previous studies [e.g. 1, 3-5]. These findings are particularly surprising, as even previous studies using imaginary scenarios found strongly increased pitch and exaggerated pitch contours in IDS [e.g. 4]. To keep our methods consistent with those of Uther *et al.* [6] we only investigated mean pitch of the vowel sound in the target words. Previous research mostly derived mean pitch values from whole phrases, sentences and words, and this methodological difference could potentially be responsible for our dissimilar findings. However, because the increased pitch in IDS has been associated with high emotional affect [e.g. 1, 3], it seems possible that the lack of a real infant in the imaginary interactions prevented the speakers from using this affective-acoustic feature in their voice as suggested by Schaeffler *et al.* [9]. Since Uther *et al.* [6] also found increased pitch in the IDS vowel sounds, we suspect that this may be one of the acoustic features that are difficult to replicate without feedback from a real infant.

In our sample, IDS was consistently rated higher for positive vocal affect than either adult condition. Although these findings follow the same trend as the natural speech samples [6], the difference in ratings between IDS and the adult conditions was a lot smaller than in the natural speech samples. It may be that, although only slightly higher than ADS, the increased mean pitch in IDS played an important part in the raters’ decision to rate the IDS samples higher. Another possibility is that the speakers in the imaginary IDS still used exaggerated pitch contours, albeit to a lesser degree than the mothers in the natural speech interactions. We are currently in the process of investigating this possibility by using new methods we recently evaluated [11] to compare the pitch contours in our imaginary sample with those of the original mothers.

We observed the same trend for the difference of emotional affect between the two adult conditions as Uther *et al.* [6], where FDS was rated to contain less positive vocal affect than ADS. One of the reasons for this finding could be an intentional reduction in speech rate by the speakers in FDS compared to ADS. This observation is consistent with Biersack *et al.*’s [7] earlier study based on imaginary speech, in which speakers modified their speech rate for an imaginary foreigner. It may be that a slower speech rate is the most accessible and obvious modification to enhance intelligibility for a foreign speech partner, and speakers may therefore use it in natural as well as in imaginary situations. It was not possible to test this possibility, as speech rate was not investigated in Uther *et al.*’s [6] study. However, we did ask our participants how they perceived their voices to have changed in each interaction, and the most common response for FDS was that they indeed reduced their speech rate. We are currently in the process of investigating the speech rate in both samples.

## 5. Conclusions

Our findings generally do not approximate those of Uther *et al.* [6] except for ratings of affect. In particular, certain crucial modifications such as hyperarticulation are not encountered in imaginary IDS. Our results therefore suggest that the use of imagined partners in speech research is probably only valid as a first step before following up with investigations using real interactions. Since hyperarticulation is thought to be an unconscious modification in IDS [e.g. 5, 6] we suspect that such unconscious modifications are those most likely to be lost. As such, the two-way dynamic feedback between speaker and listener is probably fundamental in the process of generating appropriate speech modifications. The importance of providing genuine speech partners for speech research therefore should not be underestimated if results of such studies are required to be generalisable to real-world situations.

Additional material:

<http://userweb.port.ac.uk/~knollm/additional/interspeech.htm>

## 6. Acknowledgements

We thank Prof Alan Costall for supervision and comments on this project. We also thank Dr Maria Uther, who supervised M. Knoll during data collection of the speech samples for her dissertation, published as Uther *et al.* [6]. This research was supported by a grant from the Economic and Social Research Council (ESRC) to M. Knoll.

## 7. References

- [1] Kitamura, C. and Burnham, D. “Pitch and communicative intent in mother’s speech: adjustments for age and sex in the first year”, *Infancy*, 4, 85-110, 2003.
- [2] Kuhl, P. K. “Early language acquisition: cracking the speech code”, *Nature*, 5, 831-842, 2004.
- [3] Fernald, A., Simon, T. “Expanded intonation contours in mother’s speech to newborns”, *Dev. Psychology*, 20, 104-113, 1984.
- [4] Trainor, L., Austin, C., and Desjardins, R. “Is infant-directed speech prosody a result of the vocal expression of emotion?”, *Psych. Science*, 11(3), 188-195, 2000.
- [5] Burnham, D., Kitamura, C., and Vollmer-Conna, U. “What’s new pussycat? On talking to babies and animals”, *Science*, 296, 1095, 2002.
- [6] Uther, M., Knoll, M., and Burnham, D. “Do you speak E-n-g-l-i-s-h? A comparison of foreigner- and infant-directed speech”, *Speech Communication*, 49, 2-7, 2007.
- [7] Biersack, S., Kempe, V., and Knapton, L. “Fine-Tuning Speech Registers: A Comparison of the Prosodic Features of Child-Directed and Foreigner-Directed Speech”, *Proc. 9th Euro. Conf. Speech Comm. & Techn.*, Lisbon, 2005.
- [8] Papousek, M. and Hwang, S-F. “Tone and intonation in Mandarin babytalk to presyllabic infants: Comparison with registers of adult conversation and foreign language instruction”, *Appl. Psycholing.*, 12, 481-504, 1991.
- [9] Schaeffler, F., Kempe, V., and Biersack, S. “Comparing vocal parameters in spontaneous and posed child-directed speech”, *Proc. 3<sup>rd</sup> Speech Prosody*, Dresden 688-691, 2006.
- [10] Boersma, P. and Weenink, D. “Praat: doing phonetics by computer (Version 4.1.19)”, Retrieved 06/2004 from <http://www.praat.org/>, 2004.
- [11] Knoll, M., Uther, M., MacLeod, N. O’Neill, M. & Walsh, S. “Emotional, linguistic or cute? The function of pitch contours in infant- and foreigner-directed speech”, *Proc. 3<sup>rd</sup> Speech Prosody*, Dresden 165-168, 2006.