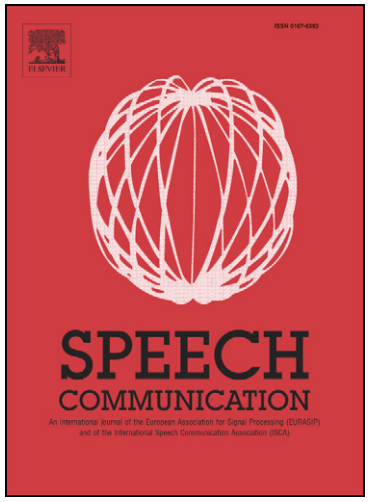
Are actresses better simulators than female students? The effects of simulation on prosodic modifications of infant- and foreigner-directed speech

Monja Knoll, Lisa Scharrer, Alan Costall



Are actresses better simulators than female students? The effects of simulation on prosodic modifications of infant- and foreigner-directed speech

Monja Knoll, Lisa Scharrer, Alan Costall

PII: S0167-6393(08)00147-7
DOI: 10.1016/j.specom.2008.10.001
Reference: SPECOM 1756

To appear in: *Speech Communication*

Received Date: 29 April 2008
Revised Date: 7 July 2008
Accepted Date: 6 October 2008

Please cite this article as: Knoll, M., Scharrer, L., Costall, A., Are actresses better simulators than female students? The effects of simulation on prosodic modifications of infant- and foreigner-directed speech, *Speech Communication* (2008), doi: 10.1016/j.specom.2008.10.001

Running head: Analyses of infant- and foreigner-directed speech

Are actresses better simulators than female students? The effects of simulation on prosodic modifications of infant- and foreigner-directed speech[1]

Monja Knoll[a*], Lisa Scharrer[a,b], Alan Costall[a]

[a]Department of Psychology, University of Portsmouth, King Henry Building, King Henry I Street, Portsmouth, PO1 2DY, United Kingdom. Email: monja.knoll@port.ac.uk; lisa.scharrer@port.ac.uk; alan.costall@port.ac.uk

[b]Psychologisches Institut, Universitaet Heidelberg, Hauptstrasse 47-51, 69117 Heidelberg, Germany. Email: Lisa.Scharrer@urz.uni-heidelberg.de

*Corresponding author: Department of Psychology, University of Portsmouth, King Henry Building, King Henry 1 Street, Portsmouth, PO1 2DY, United Kingdom. Email: monja.knoll@port.ac.uk. Tel: +44 (0) 23 9284 6317. Fax: +44 (0) 23 9284 6300

[1] Part of this work has been presented at the 8th Interspeech conference in Antwerp, Belgium (08/2007).

ABSTRACT

Previous research has used simulated interactions to investigate emotional and linguistic speech phenomena. Here, we evaluate the use of these simulated interactions by comparing speech addressed to imaginary speech partners produced by psychology students and actresses, to an existing study of natural speech addressed to genuine interaction partners. Simulated infant- (IDS), foreigner- (FDS) and adult-directed speech (ADS) was obtained from ten female students and ten female actresses. These samples were acoustically analysed and rated for positive vocal affect. Our results for affect for actresses and student speakers are consistent with previous findings using natural interactions, with IDS rated higher in positive affect than ADS/FDS, and ADS rated higher than FDS. In contrast to natural speech, acoustic analyses of IDS produced by student speakers revealed a smaller vowel space than ADS/FDS, with no significant difference between those adult conditions. In contrast to natural speech (IDS > ADS/FDS), the mean F_0 of IDS was significantly higher than ADS, but not than FDS. Acoustic analyses of actress speech were more similar to natural speech, with IDS vowel space significantly larger than ADS vowel space, and with the mean FDS vowel space positioned between these two conditions. IDS mean F_0 of the actress speakers was significantly higher than both ADS and FDS. These results indicate that training plays an important role in eliciting natural-like speech in simulated interactions, and that participants without training are less successful in reproducing such speech. Speech obtained from simulated interactions should therefore be used with caution, and level of experience and training of the speakers should be taken into account.

Index terms: IDS, simulated speech, actresses, hyperarticulation

1. Introduction

An underlying dilemma in research is to remain proximate to the actual phenomenon under investigation, while at the same time achieving effective experimental control. In linguistic and emotional speech studies, researchers have attempted to resolve this dilemma by using simulated speech to obtain standardised speech samples for comparison across different experimental conditions. In such studies, speakers are provided with specific instructions to address an imaginary audience or perform a particular scenario (e.g., Biersack et al., 2005; Papousek and Hwang, 1991; Picheny et al., 1985; Smiljanić and Bradlow, 2005; Trainor et al., 2000). However, the effects of such simulated interactions, for instance on prosodic modifications, are not well understood and merit further examination.

One area in which such methodologies have recently been used is research into prosody of speech directed to infants (IDS) and relevant comparison groups. These speech recipient groups will form the focus of the present investigation. IDS is acoustically and phonetically modified compared to adult-directed speech (ADS) (e.g. Kitamura and Burnham, 2003; Kuhl, 2004; Kuhl et al., 1997; Stern et al., 1983), and is characterised by exaggerated pitch contours, increased mean pitch, hyperarticulated vowels and high emotional affect (e.g. Burnham et al., 2002; Fernald and Simon, 1984; Stern et al., 1983; Trainor et al., 2000). Acoustic modifications in IDS may have linguistic, emotional-affective and attentional functions (e.g. Fernald and Kuhl, 1987; Fernald and Simon, 1984; Kitamura and Burnham, 1998, 2003; Kuhl et al., 1997; Papousek et al., 1991). A growing body of research has attempted to separate these functions by comparing IDS with speech addressed to other linguistic (e.g. foreigners (FDS): Papousek and Hwang, 1991; Uther et al., 2007) or emotional groups (e.g. pets

or partners: Burnham et al., 2002; Trainor et al., 2000), and a variety of different methodologies has been used in these investigations. For instance, IDS has been compared to ADS using student speakers and imagined interaction partners (simulated speech; e.g. Biersack et al., 2005; Papousek and Hwang, 1991) or using mothers in natural interactions with real interaction partners (Burnham et al., 2002; Englund and Behne, 2005; Uther et al., 2007).

Simulated speech, which usually takes place in a laboratory context, offers a greater potential than natural speech to control factors such as speaker variability, room acoustics, speech content and consistency of recordings (e.g. Englund and Behne, 2005). Using speakers' simulated speech to relevant groups is also 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. However, resulting interactions of simulated speech might be contrived and unnatural. For instance, a person's ability to simulate speech addressed to certain listener groups might depend both on their previous exposure to these groups, and their 'acting' ability. The advantage of natural interactions is that they are genuine rather than contrived.

In view of these considerations, there is clearly a need to evaluate natural and simulated comparative IDS-based research to determine whether the two differ and in what way. For instance, two recent studies (Biersack et al., 2005; Uther et al., 2007) used these different methodologies to investigate differences between IDS (or child-directed speech), FDS and speech addressed to a British adult control group (ADS). Biersack et al. (2005) used simulated speech of student speakers, whereas Uther et al. (2007) used natural interactions with mothers as speakers. With regards to fundamental frequency (F_0), both studies reported similar findings in that IDS

achieved higher F_0 measures than the adult conditions. However, because Biersack et al.'s (2005) study reported findings on speech rate and F_0 range modifications, and Uther et al.'s (2007) study concentrated on examining hyperarticulation of vowels (vowel space expansion due to shifts in Formant 1 (F1) and Formant 2 (F2), see Kuhl et al. 1997) and affective measures, the two studies are not directly comparable.

Two studies that attempted to compare natural and simulated interactions were carried out by Fernald and Simon (1984) and Schaeffler et al. (2006). Fernald and Simon (1984) found that F_0 contours in natural IDS were more exaggerated than F_0 contours in simulated IDS (both of which consisted of 'free speech' produced by mothers), but that simulated IDS presented more exaggerated F_0 contours than ADS. Both interaction types used the same mothers, who had previous experience with their infants, and should therefore be better at producing simulated IDS than for instance a student speaker without such experience. Nonetheless, these authors were unable to replicate the occurrence of increased F_0 in simulated IDS. In contrast, Schaeffler et al. (2006) found that child-directed speech with a real interlocutor (produced by mothers) differed only in terms of shimmer and harmonics-to-noise ratio from *simulated* child-directed speech (without interlocutor, produced by non-mothers), but not in terms of frequencies (F_0, F1, F2). However, compared to Fernald and Simon's (1984) 'free speech' study, Schaeffler et al. (2006) used target sentences and booklets for both their interactions, which are not entirely spontaneous.

Although both studies represent important steps in comparing simulated speech and speech directed to a real interlocutor, it remains unclear to what extent simulated interactions are comparable to genuine natural speech interactions. An understanding of the similarities and differences of these two approaches is crucial in determining whether findings of simulated studies are close approximations of natural interactions.

If simulated and natural speech are comparable, then other comparative speech types (e.g. hearing impaired speech) could be investigated using simulated interactions, particularly considering the investment in time and effort of natural interactions. We aimed to investigate this possibility by comparing results from an earlier study of natural IDS, FDS and ADS produced by mothers (Uther et al., 2007; see section 1.1 for summary) with recordings of simulated IDS, FDS and ADS by female student speakers. Student speakers were chosen because they constitute the most easily available speaker group in psychological research, and have been previously used in simulated speech research (e.g. Biersack et al., 2005; Papousek and Hwang, 1991).

However, students may vary in their ability to simulate speech without a speech partner, or may even be unable to convincingly perform such a task. As a speaker group, they might perceive the situation as contrived because they are required to carry out a task that might be unfamiliar and uncomfortable particularly if they are not used to role-play. By comparison, a group that may be better suited for such a task are actresses that have undergone professional training, and should thus be more used to role-play and simulated interactions (e.g. use of bluescreen or telephone conversations without interaction partners). Recordings of actors' speech have already been extensively and successfully used in the area of vocal emotion (e.g. Banse and Scherer, 1996; Kramer, 1964; Scherer, 2003; Scherer et al., 2001; Scherer and Ellgring, 2007; Wallbott and Scherer, 1986). Although a study by Viscovich et al. (2003) found that results of affective prosody produced by actors may be generalisable to people without professional training, actors are generally considered to be better at portraying different basic emotions than lay people (Scherer, 2003). It is thus reasonable to expect that actresses, due to their professional training, may be better at imagining talking to infants and foreigners that are not present, than psychology

students without such training. We intended to assess this possibility by comparing actresses' simulated speech samples with those of the students and the natural data set of Uther et al. (2007).

1.1 Summary of Uther et al. (2007)

To provide the reader with a better understanding of the comparative natural data set (Uther et al., 2007), we will present a summary of this study here. The Uther et al. (2007) study attempted to clarify the functional role of IDS by investigating the independence of specific acoustic components (mean F_0, vowel hyperarticulation and duration, affect) across IDS, ADS and FDS. Speakers 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). All interactions were recorded in the mothers' houses, and mothers were left alone with their infant or the confederate during the 5 to 10 minute interactions. Mothers were provided with a toy sheep, shark and shoe to elicit specific target words, but the interactions were otherwise natural and spontaneous. These target words were chosen based on previous research (Andruski and Kuhl, 1996; Burnham et al., 2002), to elicit the three corner vowels /a/, /i/, /u/ for the detection of vowel hyperarticulation and to control for differing F1/F2 values due to co-articulation (the sound 'sh' preceded each corner vowel).

Uther et al. (2007) found IDS and FDS to be characterised by vowel hyperarticulation compared to ADS. However, ADS and FDS exhibited significantly lower mean F_0 and vowel duration than IDS, with no significant differences between the adult conditions. IDS also obtained higher ratings of emotional affect compared to the adult conditions, while FDS contained less positive vocal affect than ADS. The authors concluded that, consistent with Kuhl (e.g. Kuhl et al., 1997; Kuhl, 2004),

vowel hyperarticulation is a didactic device of IDS, independent of F_0 values and affect, and suggested that vowel hyperarticulation may serve a similar purpose in FDS.

1.2 Aims of the present study

The aim of the present study was to investigate similarities and differences in prosodic modifications in IDS and FDS resulting from the use of different methodologies (simulated versus natural interactions) and speaker populations (students versus actresses). Specifically, we aimed to replicate the linguistic and affective results of the earlier natural IDS, FDS and ADS data set of Uther et al. (2007) with student and actress speakers' simulated speech. To facilitate comparison, the analyses of the present study are restricted to the same acoustic and affective measurements used by Uther et al. (2007).

If simulated speech is a close approximation of natural speech, and based on the findings of Uther et al. (2007), we hypothesise that vowel space (hyperarticulation) of both speaker groups in IDS and FDS will be greater relative to their vowel space in ADS, with no difference between IDS and FDS. Secondly, we hypothesise that the speakers' mean F_0 values and vowel duration will be greater in IDS relative to ADS and FDS, with no difference between the adult conditions. Thirdly, affective ratings of positive vocal affect of IDS will be higher relative to ADS and FDS, with ADS achieving higher ratings than FDS. However, if role-play experience and the ability to act are important for producing simulated speech samples that are viable for psychological research, as indicated by Scherer (2003), actresses would be expected to perform more similarly to the mothers' natural speech than student speakers. In that case, we hypothesise that actress speakers will produce the same pattern of prosodic

modification results for IDS, FDS and ADS as described above, with student speakers differing in at least one of these measures.

2. Method

2.1 Participants

2.1.1 Speakers

The speakers of the student interactions consisted of 10 females with a mean age of 22.9 years (*sd* 8.84) recruited from the student population of the University of Portsmouth. None of these speakers had previous acting experience. The speakers of the actress interactions consisted of 10 professional actresses with a mean age of 28.4 years (*sd* 13.06) recruited from theatres and acting schools in London. All actresses had undergone professional acting training (including voice coaching) and had been acting for more than a year (mean 5.6 years) with most of their acting work carried out in the theatre. Each of the speakers was asked to indicate their previous exposure to each of the speech groups (Likert-scale; 1 (not much exposure) to 5 (a lot of exposure)), the type of exposure, the difficulty of the task and their frustration with it, and which speech recipient group they perceived as the easiest and the most difficult. Furthermore, we collected information about how they perceived their voice to have changed, and their prediction of how other people would rate their voice on positive affect in each of the speech groups. 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 different participants for both student and actress imaginary interactions, as both sets of interactions were not carried out simultaneously. We recruited 40 students in total (20 for the student interactions; 20 for the actress interactions) 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 years (*sd* 10.37) for the ratings of the student interactions. For the ratings of the actress interactions, participants consisted of 6 males and 14 females with a mean age of 26.25 years (*sd* 10.12). None of the participants was hearing impaired, and all were British citizens. These participants were required to rate low-pass filtered speech samples of the interactions for vocal affect. Inter-rater reliability was found to be high (student interactions: reliability coefficient $\alpha = .81$; actress interactions: reliability coefficient $\alpha = .88$).

2.2 Acoustic analyses and filtering

A total of 346 words were analysed for the student interactions, and 417 words were analysed for the actress interactions (compared to 496 words in Uther et al., 2007). Acoustic analyses were carried out using Praat 4.5.16 (Boersma and Weenink, 2006) and centred on the corner vowels /a/, /i/ and /u/ of the target words shark, sheep and shoe. Vowel sounds were firstly extracted using Soundforge 6.0, and the resulting sound samples were then analysed for mean F_0, vowel duration, and F1 and F2 values (taken at the midpoint of the vowel). Mean F1/F2 values were then used as x/y coordinates to plot bivariate vowel triangles for /a/, /i/ and /u/ for each speaker's production of IDS, ADS and FDS, from which vowel triangle area was calculated to detect 'hyperarticulation' of the vowels. Plotting and calculation of the vowel space

was carried out in Autocad 2000 (a software application for 2D and 3D drafting, design and modelling). The procedures used were identical to those of Uther et al. (2007) and follow those of previous studies (Burnham et al., 2002; Kuhl et al., 1997).

For the affective analysis, approximately 30 seconds of each speech sample were low-pass filtered (Hann window) 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. (2007) who also employed a 1000 Hz filter. Otherwise, the speech samples received no further modification. The 30-second speech samples were selected from the same part of the recording for each of the speaker's interactions (location on tape at approximately the 0.20 to 0.50 second mark). The first 20 seconds of the interactions were disregarded as we assumed that speakers might speak more fluently and natural later on the recording. 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. Burnham et al., 2002; Kitamura and Burnham, 2003).

2.3 Procedure

2.3.1 Speech samples

The recordings were carried out in the laboratory with digital recording equipment. All speakers were required to produce simulated speech directed to an imagined infant, a British and a foreign adult in separate interactions. For the infant interaction, the speakers were instructed to imagine that they were talking to a close family member (ideally to their own child). However, we provided no example or idea of what IDS should sound like. For the British and foreign adult interactions, speakers were instructed to imagine talking to a female stranger in her early twenties. In the

foreign interactions they were instructed to imagine that the person was a female Chinese exchange student who had been living in the UK for less than two months, and that they might encounter some communication problems. It is important to note here that we did not explicitly instruct our participants to produce clear speech in either of these interactions. To keep these interactions consistent with Uther et al. (2007) 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 to 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. Similar to the mothers in Uther et al. (2007), the speakers were left alone during the interaction. The order of the interactions for each speaker was counterbalanced, and each interaction lasted for approximately three minutes. Speakers were given 10 minutes between each interaction to prepare for the next interaction.

2.3.2 Affective ratings

Raters listened to 30 low-pass filtered speech samples (10 each in IDS, FDS and ADS for both imaginary student and actress interactions). Using a five point Likert-scale (1 (not at all) to 5 (extremely)), participants were instructed to rate subjectively how much positive vocal affect they could detect in each 30-second speech sample. The presentation order of the speech samples was randomised (using Microsoft Office Excel 2003). This order was presented to the first half of the participants, and then reversed for the remainder of the participants. Each participant was supplied with two test trials of the filtered speech to familiarise themselves with the filtered sound.

3. Results

The data for the acoustic and affective analyses were subjected to 2 x 3 mixed measures ANOVAs with the speakers (actresses and students) as between-subjects factor (conditions) and the three speech recipient groups (IDS, ADS and FDS) as the within-subjects factor. Differences between groups were further explored with planned contrasts. Where any deviations from the assumptions of ANOVA were found, the ANOVA analyses were complemented by non-parametric tests. It should be noted here that ADS represented a baseline condition in the original study, with FDS as a linguistic comparison group to IDS.

3.1 Previous exposure to groups

We investigated whether actress and student speakers differed in self-reports of how much exposure they had to infants and foreigners. The means indicated that actresses rated themselves as having had more exposure to infants (mean 3.1, *sd* 1.2) and foreigners (mean 3.0, *sd* 1.2) than student speakers (IDS mean 2.9, *sd* 1.4; FDS mean 2.7, *sd* 1.5), and that overall both speaker groups rated themselves to have had more exposure to infants than to foreigners. However, these differences were not statistically significant, although it should be noted that these are self-reports and a reliable quantification of previous exposure would be difficult to achieve. In an attempt to further define whether the speakers differed in the type of exposure they reported, we divided their qualitative responses into four categories for IDS (casual (e.g. on the street, not much contact), friends (friends of the participants who had an infant), family (e.g. niece, nephew), own child (included older children)) and FDS

(casual (e.g. on the street; not much contact), work (worked regularly with foreigners), friends (friends were foreigners), family (e.g. stepfather)). Distribution of the speakers in each of these categories can be found in Table 1. Fisher's exact test indicated that actresses and students did not report significantly different exposure to these categories. As these categories were not necessarily progressive, we collapsed them further into 'casual' (indicating little contact) and 'regular' (indicating regular contact with IDS and FDS, and encompassed the remaining categories for these speech recipient groups). However, we found no statistical difference between these categories for the actresses and the students.

Table 1 about here

There was a difference between student speakers and actresses in their perception of the difficulty, frustration and discomfort of the task. Student speakers (Median = 2.5, 2.5, 3.0 respectively) rated the task as more frustrating (U = 14.00, p < .002, r =-.648), difficult (U = 24.00, p =.025, r = .459) and uncomfortable (U = 8.00, p < .001, r = -.741) than the actresses (Median = 1.0, 2.0, 1.0 respectively). Eight of the ten student speakers reported that they found IDS the easiest condition (two reported ADS), and all of the student speakers reported FDS as the most difficult condition. Similarly eight of the ten actress speakers reported IDS as the easiest condition (two reported ADS), and the same number of actresses reported FDS as the most difficult condition (two reported ADS).

3.2 Acoustic analyses

3.2.1 F1/F2 vowel hyperarticulation

Mean formant values for the three corner vowels and the resulting vowel triangles for each speech recipient group for students and actresses can be found in Figure 1. We found a significant main effect of condition for vowel space ($F_{(1, 18)}$ = 14.136, p = .001, η = .440), with the actress speakers producing greater vowel spaces across the three speech recipient groups than the student speakers (see Fig. 1 and 2). We did not find a significant main effect of speech recipient groups, but a significant interaction between the conditions and the three speech recipient groups ($F_{(2, 36)}$ = 7.852, p = .001, η = .304), which is due to an interaction between IDS and ADS ($F_{(1, 18)}$ = 19.889, p = .001, η = .525).

Figure 1 about here

We explored these interaction effects further with planned contrasts for each of the groups. In the student imaginary interactions, IDS vowel space was significantly smaller than both ADS ($F_{(1, 9)}$ = 6.196, p = .034, η = .408) and FDS ($F_{(1, 9)}$ = 8.982, p = .015, η = .499; Fig. 2), with no significant difference between the adult conditions. In contrast, IDS vowel space in the actress interactions was significantly larger than ADS ($F_{(1, 9)}$ = 13.935, p = .005, η = .608), but no significant differences between FDS and IDS, and between FDS and ADS were found (Fig. 2).

Figure 2 about here

The above results were confirmed with a Friedman two-way analysis of ranks on the effect of speech recipient groups for students ($\chi^2_{(2)}$ = 7.800, p =.018) and for actresses ($\chi^2_{(2)}$ = 12.200, p =.001). Wilcoxon tests showed that IDS vowel space was significantly lower than FDS (z = -2.176, p = .021), and ADS (z = -2.495, p = .028), with no difference between ADS and FDS for the student speakers. ADS vowel space was significantly lower than IDS (z = -2.889, p = .001) and FDS (z = -2.565, p = .006), with no difference between IDS and FDS for the actress speakers.

3.2.2 Vowel duration

Our analyses for vowel duration revealed a significant main effect for condition ($F_{(1, 18)}$ = 8.574, p = .009, η = .323), with students achieving significantly higher vowel duration (mean 0.114 seconds, *sd* 0.024) than actresses (mean 0.091, *sd* 0.023). However, we found no significant main effect for speech recipient groups, or interaction effect between speech recipient groups and conditions. This is in contrast to Uther et al. (2007), where IDS achieved significantly higher vowel duration than either ADS or FDS with no difference between those two conditions.

3.2.3 Fundamental frequency (F_0)

We found no significant main effect of conditions, but a significant main effect of speech recipient group ($F_{(2, 36)}$ = 20.898, p < .001, η = .537). IDS achieved significantly higher F_0 values than FDS ($F_{(1, 18)}$ = 25.684, p < .001, η = .588) and ADS ($F_{(1, 18)}$ = 40.123, p < .001, η = .690), with no difference between the adult conditions (Fig. 3). However, these main effects must be interpreted in light of the significant interaction between the conditions and speech recipient groups ($F_{(2, 36)}$ = 3.910, p =

.029, η = .178). As can be seen in Figure 3, there is a difference in the two conditions due to an interaction between IDS and FDS ($F_{(1, 18)}$ = 6.251, p = .022, η = .258).

Figure 3 about here

These effects are further demonstrated by comparing the differences between the three speech recipient groups for each of the conditions. In the student interactions, IDS achieved significantly higher mean F_0 than ADS ($F_{(1, 9)}$ = 15.424, p = .003, η = .632), but there was no significant difference between IDS and FDS, or between FDS and ADS (Fig. 3). In contrast to this, actress-produced IDS mean F_0 was significantly higher than both FDS ($F_{(1, 9)}$ = 22.678, p = .001, η = .716) and ADS ($F_{(1, 9)}$ = 25.035, p = .001, η = .736), with no difference between the adult conditions. These results were identical when confirmed with Friedman two-way analysis of ranks, Wilcoxon sign rank tests and the Chi-square approximations to the Friedman statistics for each group (F-ratio comparison between the two conditions).

3.3. Affective analyses

3.3.1 Raters

There was a significant difference for ratings of positive vocal affect between the speech recipient groups ($F_{(2, 76)}$ = 62.927, p < .001, η = .623). IDS obtained significantly higher ratings of positive vocal affect than ADS ($F_{(1, .38)}$ = 61.012 p < .001, η = .616) or FDS ($F_{(1, 38)}$ = 101.411, p < .001, η = .727). ADS also achieved significantly higher ratings than FDS ($F_{(1, 38)}$ = 6.958, p = .012, η = .155). We did not find a significant main effect for condition, or a significant interaction between speech

recipient groups and conditions. As such the two speaker groups did not differ in ratings of positive vocal affect (Fig. 4).

Figure 4 about here

3.3.2 Speakers' self ratings

We also asked the speakers how they thought other people would rate their speech samples on positive vocal affect for each of the three speech types. We included this question to investigate their awareness of the effect their simulated speech might have on naïve raters, and to compare the results with the perceptual ratings of the low-pass filtered speech. Interestingly, there was no difference between actresses and student speakers, in that both groups were conscious that their speech would be rated differently depending on who they had imagined talking to ($F_{(2, 36)} = 16.251$, $p < .001$, $\eta = .474$; see Fig. 5). Both groups assumed that their speech to the infants would be rated more positively than FDS ($F_{(1, 18)} = 33.800$, $p < .001$, $\eta = .653$) and ADS ($F_{(1, 18)} = 12.898$, $p = .002$, $\eta = .417$). Interestingly, both actress and student speakers assumed that their speech to foreigners would be rated less positively than ADS ($F_{(1, 18)} = 6.050$, $p = .043$, $\eta = .209$). These results were identical when confirmed with Friedman two-way analysis of ranks, Wilcoxon sign rank tests and the Chi-square approximations to the Friedman statistics for each group (F-ratio comparison between the two conditions). These findings of the self-ratings are matched by the actual results of the naïve raters, and show that both speaker groups had some understanding of how their speech would be perceived.

Figure 5 about here

4. Discussion

The aim of our study was to investigate the validity of simulated infant- and foreigner-directed speech of student and actress speakers by comparing it to natural speech produced by mothers. Our results suggest that, while some of the acoustic and affective aspects of natural speech samples can be reproduced in simulated speech conditions, others cannot. We also found a distinct difference between student speakers and actresses in their ability to simulate the relevant speech types.

4.1 F1/F2 vowel hyperarticulation and duration

Hyperarticulation of vowels has been found to play an important part in the development of infants' vowel categories and therefore speech acquisition (e.g. Andruski and Kuhl, 1996; Kuhl et al., 1997). Not only do mothers across diverse languages use these acoustic features in interactions with their infants (e.g. Andruski et al., 1999; Kuhl et al., 1997), but it has also been found that mothers' F1/F2 vowel expansion is related to their infants' skills in speech discrimination (Liu et al., 2003). Uther et al.'s (2007) finding of the occurrence of hyperarticulation in IDS and FDS provided further support for the linguistic role of this acoustic feature in both speech groups. However, we were unable to replicate the occurrence of hyperarticulation in either IDS or FDS with our student speakers. If we consider ADS as a normal baseline form of speech, our findings are particularly interesting as the students in our study in effect *hypo*articulated in IDS, in that 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 conditions reported that their 'dialogue' was uttered in an explanatory fashion which focused on the target words (e.g. explained where the animal (shark) might be found, or the material from which it was made). 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. The speakers may have used clearer (hyperarticulated) speech in the two adult conditions as part of a more instructional 'teacher' manner of speaking.

The results for vowel hyperarticulation of the actresses followed the trend of the natural speech samples of Uther et al. (2007). Actresses clearly expanded their vowel space when they imagined talking to an infant compared to when they were talking to a British adult, which is consistent with previous research (e.g. Burnham *et al.* 2002, Kuhl et al., 1997). Interestingly, nine out of the ten actresses also expanded their vowel space when they imagined talking to the foreigner than when they were talking to the British adult. Overall, the vowel space of actresses was more expanded than that of the students or of the mothers (see Uther et al., 2007). This could be explained by the observation that actresses use exaggerated language for their work – particularly if they are carrying out a lot of theatre work – or that they overemphasise appropriate cues (e.g. Scherer, 2003). Interestingly, neither the actress nor the student speakers replicated the greater vowel duration of IDS compared to the adult conditions in the natural speech samples. This finding supports Uther et al.'s (2007) suggestion that vowel hyperarticulation and duration are independent of each other.

Nonetheless, our results show that actresses were capable of reproducing the expanded vowel space previously found in natural speech samples (e.g. Burnham et al., 2002; Kuhl et al., 1997, Uther *et al*, 2007), and previously assumed to have been

an unconscious process that might need the presence of an infant. Despite their slightly higher mean age, the actresses did not report more exposure to infants or foreigners than the students, although they reported less discomfort, frustration and difficulty with the task, which could be due to their acting experience and training. As such it can be assumed that their training/experience gave them an advantage over the students. This assumption would be consistent with Scherer's (2003) claim that actors are better than lay people at producing posed emotions that are similar to natural vocal emotions.

4.2 Fundamental frequency (F_0)

Our F_0 findings support this possibility, as our results for the actress interactions are concordant with those of the natural interactions (Uther et al., 2007). The results of F_0 for the student interactions are slightly different. Although we found increased mean F_0 in student-produced IDS compared to ADS, we found no difference between FDS and IDS, or FDS and ADS. Compared to previous research (e.g. Burnham et al., 2002; Fernald and Simon, 1984; Grieser and Kuhl, 1988; Uther et al., 2007; Werker et al., 1994) using mothers and infants, the mean difference between IDS and both adult conditions for the student speakers is lower than expected.

The student F_0 findings are surprising, as even previous studies using imaginary scenarios found significantly increased F_0 measures in IDS (or child-directed speech) compared to both adult conditions (e.g. Biersack et al., 2005; Papousek and Hwang, 1991). However, our findings are comparable to Fernald and Simon (1984), who also reported a lack of increased F_0 in their simulated IDS compared to ADS. Similar to Fernald and Simon (1984) and to keep our methods consistent with those of Uther et al. (2007), we instructed our speakers to produce their own imaginary 'free speech',

whereas previous research generally used standardised sentences or utterances (e.g. Papousek and Hwang, 1991, Schaeffler et al., 2006). This methodological difference could potentially be responsible for our dissimilar findings, as the speakers in our present sample might have been unable to produce the increased F_0 in IDS, because they were distracted by the additional task of inventing their own scenarios. We are currently in the process of investigating standardised sentence production in the same speaker population. However, because the increased F_0 in IDS has been associated with high emotional affect (e.g. Fernald and Simon, 1984; Kitamura and Burnham, 2003), it seems also possible that the lack of a real infant in the imaginary interactions impeded the full use of this affective-acoustic feature by student speakers. We suspect that the increase of F_0 may be one of the acoustic features that are difficult to replicate without previous experience, feedback from a real infant or without professional training (e.g. actresses). Future research should investigate this possibility by using real infants or toddlers as speech recipients in the students' simulated interactions.

4.3 Emotional affect

Interestingly with regards to emotional affect, the results for both actresses and student speakers were similar to those of the mothers. IDS was consistently rated higher for positive vocal affect than either adult condition. Although these findings follow the same trend as those of the natural speech samples (Uther et al., 2007) the difference in ratings between IDS and the adult conditions for the actresses and particularly for the students was smaller than in the natural speech samples. It may be that, although only slightly higher than ADS in the student interactions, the increased mean F_0 in IDS played an important role in the raters' decision to rate the IDS samples higher. This would also be consistent with the notion that voice pitch (F_0) serves as a vehicle for

the heightened affect in IDS (e.g. Fernald and Kuhl, 1987; Kitamura and Burnham, 1998). Another possibility is that the speakers in the imaginary IDS still used exaggerated F_0 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 (Eigenshape analysis; Knoll et al., 2007) to compare the F_0 contours in our imaginary samples 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. (2007), where FDS was rated to contain less positive vocal affect than ADS for both student and actress interactions. One of the reasons for this finding could be an intentional reduction in speech rate by the speakers in FDS compared to ADS. It has been found that listeners tend to evaluate slower speech rate more negatively than faster speech rate (particularly younger speakers: 20-22 years old; Stewart and Bouchard Ryan, 1982), and that they also rate speakers with a faster speech rate as more competent and social attractive than slower speakers (Street et al., 1983). It is possible that the speakers in the current study (with a similar mean age to the speakers in Stewart and Bouchard Ryan, 1982) reduced their speech rate in the imaginary foreign interaction, and that the raters used this modification as a cue for their negative evaluation of FDS.

This possibility would be consistent with Biersack et al.'s (2005) earlier study based on simulated speech, in which speakers modified (reduced) 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 alternative explanation through comparison, as speech rate was not investigated in Uther et al.'s (2007) 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 reduced their speech rate. We are currently in the process of investigating the speech rate in the three samples. Interestingly both student and actress speakers accurately predicted how people would rate their speech samples, suggesting that speakers may have an inherent understanding of how their speech is perceived.

5. Conclusions

Our results show that only some aspects of natural speech can be reproduced with simulated speech conditions, and that a successful reproduction of other acoustic aspects may depend on the experience and training of the speakers. It seems that affective modifications (quantified by raters' perception of emotional affect) can be elicited in simulated speech regardless of the speaker's background and experience. Certain crucial acoustic modifications such as hyperarticulation are not encountered in student imaginary IDS, but can be partly replicated by actresses with the necessary training. Our results firstly add further support to previous findings (Burnham et al., 2002; Uther et al., 2007) that vowel hyperarticulation in IDS and FDS occurs independent of F_0 modifications and affect. Secondly our results 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, and should depend on the experiences and training of the speakers. Since hyperarticulation is thought to be an unconscious modification in IDS (e.g. Burnham et al., 2002; Uther et al., 2007) we suspect that such unconscious modifications are those most likely to be lost (as in the case of the student speakers), or that they at least require some previous training (e.g. actresses) or explicit instructions in order to be elicited. As such, the two-way

dynamic feedback between speaker and listener and professional experience/training are probably fundamental in the process of generating appropriate speech modifications. The importance of providing either genuine speech partners or using professional speakers for speech research should therefore not be underestimated if the results of such studies are required to be generalisable to real-world situations.

6. Acknowledgements

We thank Dr Maria Uther (Brunel University, UK), Dr Stig Walsh (NHM London) and Dr Darren Van Laar (University of Portsmouth) for useful comments on the project. Our particular thanks go to David Bauckham and the 'The Bridge Theatre Training Company' for their invaluable help in recruiting the actresses used in this study. This research was supported by a grant from the Economic and Social Research Council (ESRC) to Monja Knoll.

7. References

Andruski, J.E., Kuhl, P.E., 1996. The acoustic structure of vowels in mothers' speech to infants and adults. In: Proceedings of the 4th International Conference on Spoken Language Processing (pp. 1545-1548), Philadelphia, PA.

Andruski, J.E., Kuhl, P.K., Hayashi, A., 1999. Point vowels in Japanese mothers' speech to infants and adults. Journal of Acoustic Society of America 102, 1095-1096.

Banse, R., Scherer, K.R., 1996. Acoustic profiles in vocal emotion expression. Journal of Personality and Social Psychology 70, 614-636.

Biersack, S., Kempe, V., Knapton, L., 2005. Fine-tuning speech registers: A comparison of the prosodic features of child-directed and foreigner-directed speech. In: Proceedings of the 9th European Conference on Speech Communication and Technology (pp. 2401-2405), Lisbon.

Boersma, P., Weenink, D., 2006. Praat: doing phonetics by computer (Version 4.5.16) [Computer programme]. Retrieved June 2006 from http://www.praat.org/.

Burnham, D., Kitamura, C., Vollmer-Conna, U., 2002. What's new pussycat? On talking to babies and animals. Science 296, 1435.

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.

Fernald, A., Kuhl, P., 1987. Acoustic determinants of infant preference for motherese speech. Infant Behavior and Development 10 (3), 279-293.

Fernald, A., Simon, T., 1984. Expanded intonation contours in mothers' speech to newborns. Developmental Psychology 20, 104-113.

Griesser, D.L, Kuhl, P.K., 1988. Maternal speech to infants in a tonal language: support for universal prosodic features in motherese. Developmental Psychology 24 (1), 14-20.

Kitamura, C., Burnham, D., 1998. The infant's response to vocal affect in maternal speech, in: Rovee-Collier, C. (Ed.), Advances in Infancy Research, Ablex Publishing, Norwood, NJ, pp. 221-236.

Kitamura, C., Burnham, D., 2003. Pitch and communicative intent in mother's speech: Adjustments for age and sex in the first year. Infancy 4, 85-110.

Knoll, M.A., Walsh, S.A., MacLeod, N., O'Neill, M., Uther, M., 2007. Good performers know their audience! Identification and characterisation of pitch contours in infant and foreigner-directed speech, in: MacLeod, N. (Ed.), Automated taxon recognition in systematics: Theory, approaches and applications, CRC Press/The Systematics Association, Boca Raton, Florida, pp. 299-310.

Kramer, E., 1964. Elimination of verbal cues in judgments of emotion from voice. Journal of Abnormal Social Psychology 68, 390-396.

Kuhl, P.K., 2004. Early language acquisition: Cracking the speech code. Nature Reviews Neuroscience 5, 831-843.

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 (5326), 684-686.

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

Papousek, M., Hwang, S.-F.C., 1991. Tone and intonation in Mandarin babytalk to presyllabic infants: Comparison with registers of adult conversation and foreign language instruction. Applied Psycholinguistics 12, 481–504.

Papousek, M., Papousek, H., Symmes, D., 1991. The meaning of melodies in motherese in tone and stress languages. Infant Behavior and Development 14, 415-440.

Picheny, M.A., Durlach, N.I., Braida, L.D., 1985. Speaking clearly for the hard of hearing I: intelligibility differences between clear and conversational speech. Journal of Speech and Hearing Research 28, 96-103.

Schaeffler, F., Kempe, V., Biersack, S., 2006. Comparing vocal parameters in spontaneous and posed child-directed speech. Paper presented at the 3rd Speech Prosody Congress, Dresden, Germany.

Scherer, K.R., 2003. Vocal communication of emotion: A review of research paradigms. Speech Communication 40, 227-256.

Scherer, K.R., Banse, R., Wallbott, H.G., 2001. Emotion inferences from vocal expression correlate across languages and cultures. Journal of Cross-Cultural Psychology 32, 76–92.

Scherer, K.R., Ellgring, H., 2007. Multimodal expression of emotion: Affect programs or componential appraisal patterns? Emotion 7 (1), 158-171.

Smiljanić, R., Bradlow, A.R., 2005. Production and perception of clear speech in Croatian and English. Journal of the Acoustical Society of America 118 (3), 1677-1688.

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

Stewart, M.A., Bouchard Ryan, E., 1982. Attitudes toward younger and older adult speakers: Effects of varying speech rates. Journal of Language and Social Psychology 1 (2), 91-109.

Street, R.L., Brady, R.M., Putman, W.B., 1983. The influence of speech rate stereotypes and rate similarity or listeners' evaluations of speakers. Journal of Language and Social Psychology 2 (1), 37-56.

Trainor, L.J., Austin, C.M., Desjardins, R.N., 2000. Is infant-directed speech prosody a result of the vocal expression of emotion? Psychological Science 11, 188-195.

Uther, M., Knoll, M.A., Burnham, D., 2007. Do you speak E-n-g-l-i-s-h? A comparison of foreigner- and infant-directed speech. Speech Communication 49, 2-7.

Viscovich, N., Borod, J., Pihan, H., Peery, S., Brickman, A.M., Tabert, M., Schmidt, M., Spielman, J., 2003. Acoustical analysis of posed prosodic expressions: Effects of emotion and sex. Perceptual and Motor Skills 96, 759-771.

Wallbott, H.G., Scherer, K.R., 1986. Cues and channels in emotion recognition. Journal of Personality and Social Psychology 51, 690-699.

Werker, J.F., Pegg, J.E., McLeod, P.J., 1994. A cross-language investigation of infant preference for infant-directed communication. Infant Behaviour and Development 17, 323-333.

Table captions

Table 1. *Comparison of qualitative self-reports of exposure to IDS and FDS for student and actress speakers (values represent frequency counts).*

	IDS				FDS			
	Casual	Friends	Family	Own infant	Casual	Work	Friends	Family
Students	4	1	4	1	4	1	3	2
Actresses	3	3	3	1	3	2	4	1

Figure captions

Figure 1. *Mean vowel triangles as indexed by plotted F1/F2 mean values for the three speech recipient groups (IDS, FDS and ADS) for student (left panel) and actress speakers (right panel).*

Fig. 2. *Interaction between conditions (actresses and students) and speech recipient groups (IDS, FDS and ADS) for mean vowel space expansion (hyperarticulation).*

Fig. 3. *Interaction between conditions (actresses and students) and speech recipient groups (IDS, FDS and ADS) for mean fundamental frequency. .*

Fig. 4. *Representation of positive vocal affect for imaginary student and actress IDS, ADS and FDS. Error bars represent standard error of means.*

Fig. 5. *Comparison of self-ratings of positive vocal affect for student and actress speakers. Error bars represent standard error of means.*

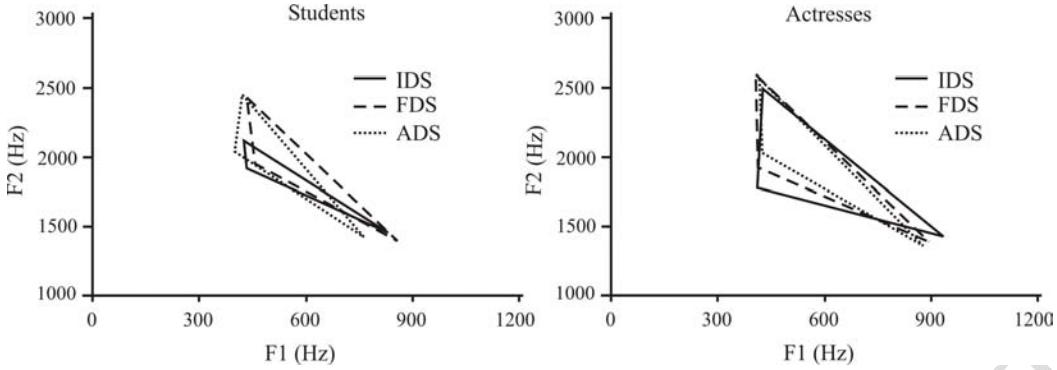


Figure 1

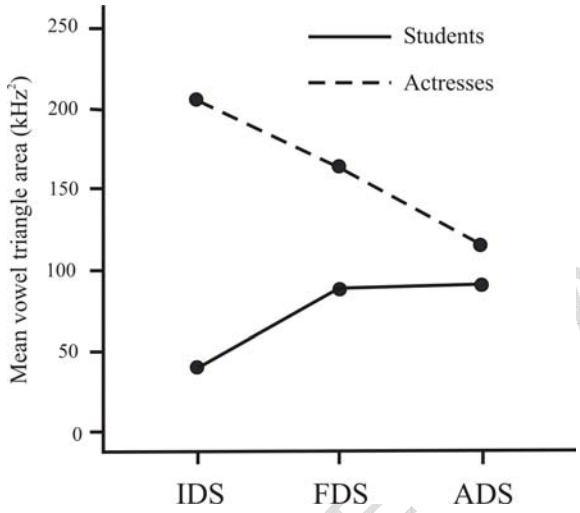


Figure 2

Figure 3

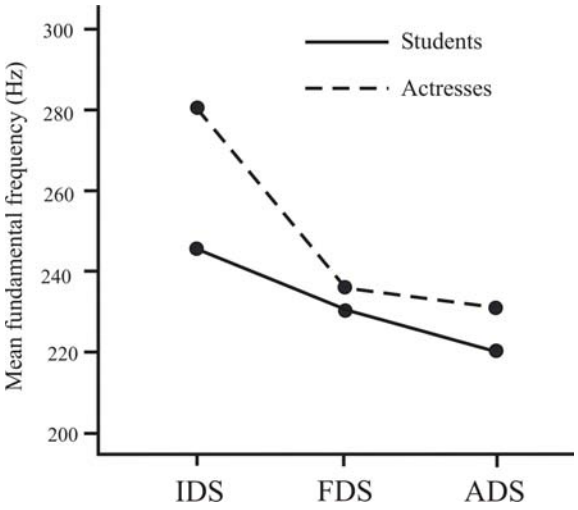


Figure 4

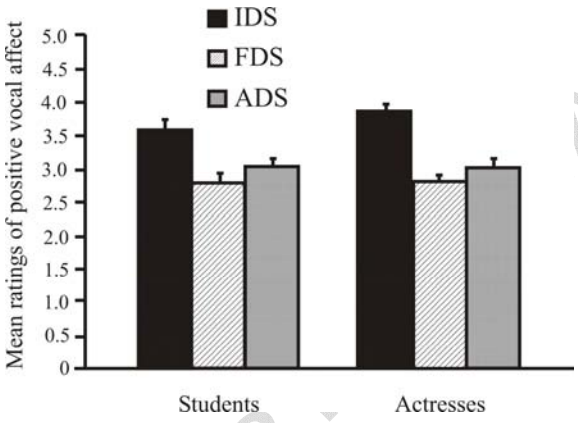


Figure 5