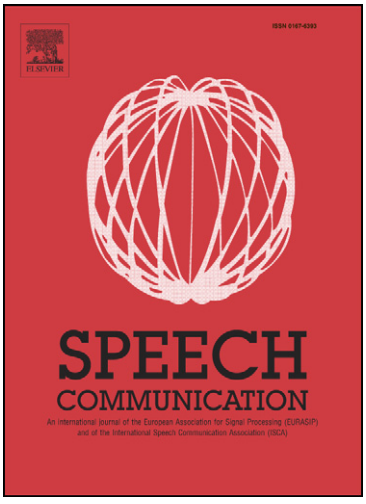
”Look at the shark”: Evaluation of student- and actress-produced standardised sentences of infant- and foreigner-directed speech

Monja Knoll, Lisa Scharrer, Alan Costall



"Look at the shark": Evaluation of student- and actress-produced standardised sentences of infant- and foreigner-directed speech

Monja Knoll, Lisa Scharrer, Alan Costall

PII: S0167-6393(10)00138-X
DOI: 10.1016/j.specom.2010.08.004
Reference: SPECOM 1918

To appear in: *Speech Communication*

Received Date: 9 March 2010
Revised Date: 10 June 2010
Accepted Date: 2 August 2010

Please cite this article as: Knoll, M., Scharrer, L., Costall, A., "Look at the shark": Evaluation of student- and actress-produced standardised sentences of infant- and foreigner-directed speech, *Speech Communication* (2010), doi: 10.1016/j.specom.2010.08.004

Running head: Evaluation of simulated sentence production

"Look at the shark": Evaluation of student- and actress-produced standardised sentences of infant- and foreigner-directed speech[1]

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

[a]School of Social Sciences, University of the West of Scotland, Paisley, PA1 2BE, United Kingdom. Email: monja.knoll@port.ac.uk.

[b]Department of Psychology, University of Muenster, Fliednerstrasse 21, 48149 Muenster, Germany. Email: Lisa.Scharrer@uni-muenster.de

[c]Department of Psychology, University of Portsmouth, King Henry Building, King Henry I Street, Portsmouth, PO1 2DY, United Kingdom. Email: alan.costall@port.ac.uk

*Corresponding author: School of Social Sciences, University of the West of Scotland, Paisley, PA1 2BE, United Kingdom.

[1] Part of this work is included in the proceedings of the 9th Interspeech conference in Brisbane, Australia (09/2008).

Abstract

Standardised sentence production is routinely used in speech research to avoid content variability typical of natural speech production. However, the validity of such standardised material is not well understood. Here, we evaluated the use of standardised sentences by comparing them to two existing, non-standardised datasets of simulated free and natural speech (the latter produced by mothers in real interactions). Standardised sentences and simulated free speech were produced by students and actresses without an interaction partner. Each dataset comprised recordings of infant- (IDS), foreigner- (FDS) and adult-directed (ADS) speech, which were analysed for mean F_0, vowel duration and hyperarticulation. Whilst students' mean F_0 pattern in standardised speech was closer to the natural speech than their previous 'simulated free speech', no difference in vowel hyperarticulation and duration patterns was found for students' standardised sentences between the three speech styles. Actresses' F_0, vowel duration and hyperarticulation patterns in standardised speech were similar to the natural speech, and a part improvement on their 'simulated free speech'. These results suggest that successful reproduction of some acoustic measures (e.g., F_0) can be achieved with standardised content regardless of the type of speaker, whereas the production of other acoustic measures (e.g., hyperarticulation) are context- and speaker-dependent.

Introduction

The acoustic characteristics of a person's voice can provide a rich source of information about, for instance, their personality and emotional state (Starkweather, 1967). However, investigations into these acoustic properties may require the separation of the verbal (semantic) from the non-verbal (acoustic) features of speech to alleviate the problem of semantic interference with acoustic and perceptual measurements. Not surprisingly, the achievement of this separation has been a frequent concern in speech research. A variety of different methodologies has been used depending on the relevant aspect of speech under study (e.g., linguistic or emotional), the type of data collection (e.g., natural conversational speech or standardised sentence production in the laboratory) and type of analysis (e.g., perceptual ratings or acoustic analysis). For instance, in emotional and personality research, a common approach has been to remove the semantic content, or to render the speech samples unintelligible by means of low-pass filtering (e.g., Cohen and Starkweather, 1961; Milmoe et al., 1967; Rogers et al., 1971; Soskin and Kaufman, 1961; Starkweather, 1956), content-splicing (e.g., Scherer, 1971; Scherer et al., 1972), speech reversal (e.g., Knower, 1941), the use of foreign speech samples (e.g., Kramer, 1964) and 'reiterant'[2] speech (e.g., Friend and Farrer, 1994).

The present study aimed to evaluate another widely used method of avoiding semantic interference, where the semantic content of speech is standardised by using identical sentences and utterances across different speakers and speech types (e.g., Anolli et al., 2000, 2002; Bradlow et al., 2003; Cheang and Pell, 2008; Ladd et al., 1985;

[2] Syllables in a word are exchanged with nonsense syllables to render the word unintelligible.

Rockwell, 2000; Scherer and Ellgring, 2007; Tischer, 1988). It should be noted that the current study is concerned with linguistic rather than emotional factors, although standardised sentences can be used to investigate either.

Standardised content

The reason that standardised sentences have been routinely used in speech research is that they facilitate perceptual and acoustic comparison across different speakers and conditions without the complication of variable content. By necessity, the use of standardised sentences precludes the collection of natural speech samples. Typically, speakers have been instructed to talk to a specific (imagined) listener (e.g., hearing impaired person or foreigner: Bradlow et al., 2003; Picheny et al., 1985; Smiljanić and Bradlow, 2005) whilst reciting the sentences, or they have been instructed to utter (read) the material in a specific emotional (e.g., happy, sad or ironic) 'tone of voice' (e.g., Anolli et al., 2000, 2002; Banse and Scherer, 1996; Bänziger and Scherer, 2005; Breitenstein et al., 2001; Cheang and Pell, 2008; Ross et al., 1973; Scherer et al., 1991). In some cases this standardised content has consisted of nonsense sentences (i.e., 'Hat sundig pron you venzy'; e.g., Banse and Scherer, 1996; Bänziger and Scherer, 2005; Scherer et al., 1991). Some studies have also provided the speakers with specific scenarios in which the target utterances were embedded in order to encourage a more 'ecologically valid'[3] voice production, and to elicit the required effect, such as emotion or irony (e.g., Cheang and Pell, 2008; Rockwell, 2000; Tischer, 1988; Trainor et al., 2000). In most cases the speaker has been asked to address an imaginary speech partner, although some studies have used an interaction partner in the form of the researcher or

[3] 'Ecologically valid' as used here refers to voice production that might be close to natural speech production.

another participant (e.g., Trainor et al., 2000; Wallbott and Scherer, 1986). Studies of this kind have utilised a variety of different speaker 'types', which include the use of professional actors (e.g., Banse and Scherer, 1996; Breitenstein et al., 2001; Laukka et al., 2005; Scherer and Ellgring, 2007; Wallbott and Scherer, 1986), lay actors (e.g., Anolli et al., 2000, 2002; Viscovich et al., 2003), students or non-professional speakers (e.g., Bradlow et al., 1996; Cheang and Pell, 2008; Papousek and Hwang, 1991; Schaeffler et al., 2006; Trainor et al., 2000) and experienced speech professionals (e.g., Ferguson and Kewley-Port, 2002; Kalikow et al., 1977; Picheny et al., 1985).

Comparative speech research

One area in which standardised sentences have recently been used is investigations into the linguistic and emotional functions of infant-directed speech (IDS; e.g., Papousek and Hwang, 1991; Schaeffler et al., 2006; Trainor et al., 2000). Here, the acoustic and perceptual properties of IDS have been compared with speech directed to linguistic (e.g., foreigners: Biersack et al., 2005; Uther et al., 2007) or emotional (e.g., partners, pets: Burnham et al., 2002; Trainor et al., 2000) comparison groups. Of course, the comparison of the semantically less diverse IDS with speech directed to adults is particularly challenging using non-standardised conversational material. Firstly, raters in perceptual studies may be able to identify the relevant speech recipient group (e.g., infants in IDS) from the semantic content alone, and their ratings may consequently be influenced by this. Secondly, in order to carry out specific acoustic analyses (e.g., formant analysis of vowel sounds), the semantic content in speech directed to different recipient groups needs to be similar. The use of standardised sentences solves both of these problems, and can also alleviate the need for an interaction partner, making data collection quicker and more

convenient (e.g., it may be difficult to recruit a foreign confederate who is available for a prolonged period of time).

However, the comparability of such material with natural speech has not been assessed. Natural speech provides speakers with the opportunity to freely express themselves without the constraint of standardised semantic content, and the resulting speech is negotiated in social interactions with the listener (*sensu* Parkinson, 1996). These aspects of 'real world' speech are eliminated in research based on standardised sentence production, as the semantic content is predefined and the resulting speech is based on imagined (simulated) interactions.

Comparison of natural and simulated interactions

Previously, we have already explored the effects of *simulated* interactions on the acoustic and perceptual properties of speech directed to different recipient groups (Knoll et al., 2009). In that earlier study, we compared natural speech samples of an existing dataset (Uther et al., 2007) of infant- (IDS), foreigner- (FDS) and British adult-directed (ADS) speech with speech directed to an imaginary interaction partner by students and actresses(Knoll et al., 2009). Student speakers were used as one comparison group because they constitute the most easily available participants in psychological research, and they have been previously used in simulated speech research (e.g., Biersack and Kempe, 2005; Biersack et al., 2005; Papousek and Hwang, 1991). Actresses were utilised as another comparison group because they should be more used to simulated role play, and recordings of actors' speech have already been extensively used in the area of vocal emotion (e.g., Banse and Scherer, 1996; Kramer, 1964; Scherer, 2003; Scherer and

Ellgring, 2007; Scherer et al., 2001; Wallbott and Scherer, 1986; Williams and Stevens, 1972).

Although in that study we found that actress-produced speech approximated (i.e., shared a pattern of similarity) with natural speech with regard to hyperarticulation[4], and was consistent with the increased mean fundamental frequency (F_0) in natural IDS (Knoll et al., 2009), we were notably unsuccessful in reproducing these acoustic features in the student simulated interactions (see Table 1 for similarities and differences across the three speech styles; Knoll et al., 2009). Interestingly, neither the actresses nor the student speakers were able to reproduce the vowel duration patterns of the natural speech samples.

< insert Table 1 approximately here >

In contrast to previous studies, which have used standardised utterances (e.g., Papousek and Hwang, 1991; Schaeffler et al., 2006; Trainor et al., 2000), we had asked our participants to produce their own speech without a script and using their own scenarios (referred to as 'simulated free speech' to distinguish it from natural speech; see Knoll et al., 2009). However, we did provide them with the same materials (three toys) upon which to centre their scenarios to elicit the same target words, and with a description of the confederates used in the natural dataset (Uther et al., 2007; but see Bradlow et al., 2007 for alternative method of eliciting controlled yet quasi-spontaneous speech – the Diapix Task). On the one hand, this approach should have resulted in speech that is closer to natural speech as the speakers were able to freely express themselves. On

[4] Hyperarticulation here refers to vowel space expansion due to shifts in formant 1 (F1) and formant 2 (F2) of the three corner vowels /a/, /i/ and /u/ (see Kuhl et al., 1997).

the other hand, the students might have been mainly concerned with producing a script relevant to IDS, FDS and ADS, which prevented them from concentrating on their primary task of reproducing *acoustic* speech modifications that are found in natural IDS and FDS. In our earlier study we suggested that it is possible that the additional task and freedom of inventing their own scenario 'dialogue' acted as a distraction to the student speakers through 'task overload' (Knoll et al., 2009). The use of standardised sentence production might thus alleviate this problem of 'task overload' for the student speakers. Although the actress speakers might have been accustomed to such 'multi-tasking' due to their acting experience and training, the use of sentences approximating actress scripts might also serve to improve the performance of the actresses for the same reasons (e.g., with regard to their vowel duration patterns, see Table 1).

Aim of the present study

This study aims to evaluate the use of standardised content in speech research by comparing speech modifications found in standardised sentences with those found in 'simulated free speech' (Knoll et al., 2009) and naturally produced speech (Uther et al., 2007) produced by three speaker groups across three different speech types (IDS, FDS and ADS). Both the standardised sentences used here and the 'simulated free speech' of Knoll et al. (2009) were produced by students and actresses in imaginary interactions, whereas the natural speech samples were produced by mothers in real interactions (Uther et al., 2007). To facilitate comparison between these three datasets, our analyses concentrate on a) mean F_0, b) vowel hyperarticulation and c) mean vowel duration, because these acoustic properties had been investigated in both the natural speech dataset (Uther et al., 2007) and the 'simulated free speech' dataset (Knoll et al., 2009). If the

differences presented in our earlier study (Knoll et al., 2009; see also Table 1) between the 'simulated free speech' and the natural dataset were due to 'task overload', we would expect student speakers to perform more similarly to the natural speech samples (e.g., similar hyperarticulation pattern: IDS = FDS > ADS; Uther et al., 2007) in a standardised sentence condition than in their earlier 'simulated free speech' condition (Knoll et al., 2009). We would also expect the actress speakers to produce acoustic modifications that are either consistent with their previous performance (see Table 1; Knoll et al., 2009), or constitute an improvement in that they are even more similar to the natural speech samples through alleviation of the previously mentioned 'task overload'.

Method

Design

We applied a 2 x 3 mixed measures design, consisting of speaker type (actresses versus students; between subjects factor) and speech recipient groups (IDS, FDS and ADS; within subjects factor). The dependent variables were the different acoustic measurements (mean F_0, mean vowel duration and mean vowel space). In addition to these acoustic measurements we also investigated whether the speech samples could be categorised into their intended respective speech recipient groups (IDS, FDS and ADS) for the actress and the student speakers by means of naïve raters. This was intended to provide an indication of how closely the samples resembled IDS, FDS or ADS perceptually, regardless of their quantitative acoustic measures.

Participants

Speakers

To facilitate comparison with the Knoll et al. (2009) study, the same students and actresses took part in the present study. Student speakers 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 the student speakers had previous acting experience. The actress group 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 at least a year (mean 5.6 years; range 1 to 16 years) with most of their acting work carried out in the theatre. For each of the speakers we collected information regarding a) their previous exposure to each of the speech recipient groups (Likert-scale: 1 (not much exposure) to 5 (a lot of exposure)), b) which speech production condition they had found easier ('simulated free speech' or standardised sentences) and c) which speech recipient group they had found easiest to address (IDS, ADS or FDS). Speakers had no reported hearing or speech impairment and 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).

Raters

The raters consisted of four males and two females with a mean age of 39 years (*sd* 13.97), and were recruited from the student and staff population of the University of Portsmouth. None of the raters was hearing impaired. These raters were required to

categorise the speech samples (sentences) into their respective speech recipient groups (IDS, FDS and ADS). Please note that the number of raters is comparable to those used in previous studies (Trainor et al., 2000).

Procedure

Speech samples

The speakers were required to imagine talking to an infant, a British adult and a foreign adult in three 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 ideally their own child). However, we did not provide speakers with an example or idea of what IDS should sound like. For both the British and foreign adult interactions, the speakers were instructed to imagine talking to a female stranger in her early twenties. In the foreign interaction, the raters were also instructed to imagine that the person was a Chinese exchange student who had been living in the UK for less than two months, and that they might encounter some communication difficulties. The order of the interactions for each speaker was counterbalanced to avoid order effects.

To elicit the target words, we supplied the speakers with three toys (a shark, a sheep and a shoe). The speakers were required to incorporate these target words in four standardised sentences ("Look at the '*target word*'" for each of the three target words in turn, and then combining all three target words in one sentence: "Look at the 'shark', the 'sheep' and the 'shoe'"). The order of production of the first three short sentences was counterbalanced for each of the speakers, but the long sentence was always produced at the end. The use of a small number of sentences per speaker is consistent with previous research incorporating standardised sentence production in IDS and ADS (e.g. Trainor et

al., 2000), and was used to avoid deterioration of speech modifications due to satiation and speaker fatigue. Furthermore, the production of the sentences for each of the three interactions had been preceded by the corresponding 'simulated free speech' condition (Knoll et al., 2009) in which the speakers constructed and invented their own scenarios centred on the target words (e.g., the IDS sentence condition would have been preceded by the IDS 'simulated free speech' condition) and in which they would have already repeated the words several times. Ideally this procedure should have resulted in a more relaxed speech production for the standard sentences condition, and is comparable to inductor texts used in emotion and irony research (e.g., Anolli et al., 2000; Banse and Scherer, 1996).

Ratings

We selected the utterance "Look at the shark, the sheep and the shoe" as material for the rating part of the experiment, because the sentence contained all three target words, and the presentation of all 240 sentences might also have led to rater fatigue. The raters were required to listen to each of the 60 sentences (30 actress and 30 student sentences), and categorise them into the relevant speech recipient group (IDS, FDS and ADS). Speech samples were between one and four seconds long, and participants were only able to listen to each speech sample once. To avoid order effects, the speech samples were firstly randomised, and half of the participants were presented with the speech samples in the reversed order. Participants were given breaks at different intervals to avoid fatigue and semantic satiation (Jakobovits, 1965).

Acoustic analyses

The procedures used here were identical to those of Knoll et al. (2009), and follow those of previous studies (Burnham et al., 2002; Kuhl et al., 1997; Uther et al., 2007). A total of 360 words were analysed (180 for each speaker group). Acoustic analyses were carried out using Praat 4.5.16, (Boersma and Weenink, 2006), and standard recommendations suggested by Praat for female voices were used. Based on previous research (Burnham et al., 2002; Uther et al., 2007) the analyses centred on the target words shark, sheep and shoe containing the corner vowels /a/, /i/ and /u/. These target words were also chosen to avoid problems of co-articulation due to different phonemes at the beginning of the words (i.e., each vowel was preceded by the phoneme /sh/). Vowel sounds were first extracted using Soundforge 6.0. The resulting sound samples were analysed for mean F_0, mean vowel duration, and F1/F2 measures (taken at the midpoint of the vowel, see Kuhl et al., 1997). Mean F1/F2 values were then used as x-y co-ordinates to plot bivariate vowel triangles for /a/, /i/ and /u/, 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.

Results

The data for the acoustic analyses (mean F_0, F1/F2 vowel area expansion and mean vowel duration) were subjected to a series of mixed measures ANOVAs, whereas the data for the ratings were analysed with chi-square analysis. To provide a complete overview, the results of the acoustic analyses we present here also contain mean values

for the 'simulated free speech' condition for the students and actresses (after Knoll et al., 2009) and for the mothers' natural speech samples (after Uther et al., 2007).

Previous exposure

As previously reported in Knoll et al. (2009), there was no significant difference between actresses' and students' self-ratings on how much exposure each group had to infants and foreigners. Similarly, there was no difference between actress and student speakers' reports on which speech recipient group they found the easiest or most difficult to address. The majority of both speaker groups reported IDS as the easiest and FDS as the most difficult condition. However, student speakers reported both tasks (standardised sentences and 'simulated free speech') as more frustrating, difficult and uncomfortable than the actress speakers.

Interestingly, eight of the ten student speakers stated that they found the 'simulated free speech' condition more difficult than the sentence condition, whereas only four of the ten actress speakers found the 'simulated free speech' condition more difficult than the sentence condition.

Hyperarticulation (F1/F2 vowel area expansion)

Mean formant values for the three corner vowels and the resulting vowel triangles for each speech recipient group can be found in Figure 1 for both the students and the actresses. Overall, the natural speech is less hyperarticulated then either the actresses' or students' speech regardless of the intended recipient or the condition (i.e. standardised sentences or 'simulated free speech'; see Figure 2). However, we were interested in comparing the pattern of hyperarticulation across the recipient types rather than the

absolute values. Uther et al. (2007) had found that IDS and FDS were significantly more hyperarticulated than ADS, with no difference between IDS and FDS. We would expect the same pattern for the students and the actresses.

There was a significant main effect of speech recipient groups ($F_{(2, 36)} = 4.903$, $p = .013$, $\eta = .214$). IDS vowel space was significantly larger than ADS vowel space ($F_{(1, 18)} = 8.778$, $p = .001$, $\eta = .328$), but not larger than FDS vowel space. FDS exhibited greater vowel space than ADS ($F_{(1, 18)} = 10.479$, $p = .005$, $\eta = .368$). A significant main effect for speaker type was also found ($F_{(1, 18)} = 14.000$, $p = .001$, $\eta = .437$), with actresses achieving a significantly greater vowel space than students (Figures 1 and 2). However, these main effects need to be interpreted in light of a significant interaction between the speakers and the speech recipient groups ($F_{(2, 36)} = 4.357$, $p = .02$, $\eta = .195$), which is due to an interaction between IDS and ADS ($F_{(1, 18)} = 7.597$, $p = .013$, $\eta = .297$), whereas the interaction between IDS and FDS only approached significance ($F_{(1, 18)} = 3.712$, $p = .07$, $\eta = .171$; Figure 2).

<insert Figure 1 about here>

To investigate these interaction effects further, planned contrasts for each of the speaker conditions were carried out. For the actresses, IDS vowel space was significantly greater than ADS vowel space ($F_{(1, 9)} = 10.989$, $p = .009$, $\eta = .550$), with no difference between IDS and FDS. FDS was significantly greater than ADS vowel space ($F_{(1, 9)} = 8.468$, $p = .017$, $\eta = .485$). In contrast, there was no significant difference between the speech recipient groups for the student interactions (Figure 2, right panel). The data for

the vowel space did not fully meet the assumption of homogeneity of variance between the students and the actresses. However, the results of the ANOVA were confirmed with a Friedman two-way analysis of ranks on the effect of speech recipient groups for students ($\chi^2_{(2)} = 3.800$, *ns*) and for actresses ($\chi^2_{(2)} = 14.600$, $p < .001$). Wilcoxon tests showed that actress-produced ADS vowel space was significantly smaller than IDS ($z = -2.803$, $p = .001$) and FDS ($z = -2.293$, $p = .01$), with no difference between IDS and FDS. Overall, the vowel space results for the actresses are similar to the results of the natural speech samples (Uther et al., 2007) and the actresses' earlier 'simulated free speech' condition (Knoll et al., 2009). In contrast, the results for the students are neither similar to their earlier 'simulated free speech' condition nor the natural speech samples (see also Figure 2).

<insert Figure 2 about here>

Vowel duration

There was no significant interaction between the speakers and the speech recipient groups for vowel duration. However, a significant main effect of speakers was observed ($F_{(1, 18)} = 4.920$, $p = .04$, $\eta = .215$), with actresses producing significantly longer vowels than students (Figure 3). There was also a significant main effect of speech recipient groups ($F_{(2, 36)} = 7.937$, $p = .001$, $\eta = .360$), with IDS exhibiting significantly longer mean vowel duration than FDS ($F_{(1, 18)} = 4.925$, $p = .04$, $\eta = .215$) and ADS ($F_{(1, 18)} = 14.639$, $p = .001$, $\eta = .449$). No significant difference was found between ADS and FDS mean vowel duration.

To investigate whether these differences would remain constant in each of the speaker groups, planned contrasts were conducted between the speech recipient groups for each of the conditions. Consistent with their earlier performance in the 'simulated free speech' condition (Knoll et al., 2009), there was no significant difference for mean vowel duration between student-produced IDS, FDS and ADS, but is in contrast to the natural speech samples (Uther et al., 2007; see Figure 3). However, a significant difference was found between mean vowel duration for actresses ($F_{(2, 18)} = 9.320$, $p = .002$, $\eta = .509$), with IDS exhibiting longer vowel duration than FDS ($F_{(1, 9)} = 6.947$, $p = .027$, $\eta = .436$) and ADS ($F_{(1, 9)} = 15.729$, $p = .003$, $\eta = .636$; Table 2), with no difference between the latter two. The results for the actresses were thus comparable to the natural speech samples (Uther et al., 2007), which constituted an improvement from their earlier 'simulated free speech' performance, where no difference in vowel duration between IDS, FDS and ADS was found (Knoll et al., 2009; Figure 3).

<insert Figure 3 about here>

Fundamental frequency (F_0)

There was no significant interaction for F_0 between the speakers and speech recipient groups, and no significant main effect for speakers. However, a significant main effect of speech recipient groups was observed ($F_{(2, 36)} = 21.302$, $p < .001$, $\eta = .542$), with IDS mean F_0 being significantly higher than ADS ($F_{(1, 18)} = 29.847$, $p = .001$, $\eta = .624$) and FDS ($F_{(1, 18)} = 21.459$, $p < .001$, $\eta = .544$). There was no difference between the adult groups (ADS and FDS; see Figure 4).

Planned contrasts for each of the speaker conditions showed that these differences between the speech groups were consistent for the actresses and students. Mean F_0 in IDS was significantly higher than in ADS (students: $F_{(1, 9)} = 7.406$, $p = .024$, $\eta = .451$; actresses: $F_{(1, 9)} = 17.252$, $p = .002$, $\eta = .657$) and FDS (students: $F_{(1, 9)} = 10.358$, $p = .011$, $\eta = .515$; actresses: $F_{(1, 9)} = 8.134$, $p = .019$, $\eta = .451$), while no difference between the adult conditions was found (see Figure 4). Overall, the pattern of F_0 results for the student speakers were consistent with those of the natural speech samples (Uther et al., 2007), and they also constituted an improvement compared to the students' previous 'simulated free speech' condition (Knoll et al., 2009). The results for the actress speakers remained consistent with their previous performance in the 'simulated free speech' condition (Knoll et al., 2009), which were already close to the natural speech samples (Uther et al., 2007; see also Figure 4).

<insert Figure 4 about here>

Ratings

The ratings data were analysed with standard frequency statistics (3 x 3 chi-square contingency table). Separate chi-square analyses for each of the raters (divided into actresses and students) were conducted. The raters' categorisations followed a similar pattern. In order to allow overall analyses and comparisons, and given that the patterns of association within the contingency tables were similar, the chi-square values for each of the individual raters were summated (see Johnson and Kotz, 1970a, p. 167). Chi-square results for the categorisation of the speech recipient groups separated into actress and

student speech samples for each of the raters can be found in Table 2[5]. Rated speech recipient group variables were found to be associated with the true intended speech recipient group variables for actresses (cumulative $\chi^2 = 154.993$, df = 24, $p < 0.01$) and for students (cumulative $\chi^2 = 66.68$, df = 24, $p < 0.01$).

To compare whether the strength of the relationship between the variables differed significantly between the actresses and students, we calculated the F-value from the ratio of their respective chi-square/degrees of freedom values (see Johnson and Kotz, 1970b, p. 75 for relationship between the chi-square and F-distribution). Chi-square ratio comparison ($F_{(24, 24)} = 2.234$, $p < 0.01$) between the two speaker groups showed that this association was significantly stronger for the actresses (mean contingency co-efficient .678) than for the students (mean contingency co-efficient .510; see Table 2).

<insert Table 2 about here>

Overall, the majority of the rated speech samples were correctly identified for each of their respective speech recipient groups for both actress and student speech samples (see Table 3 for summary of raters' categorisations). Interestingly, actress-produced IDS and FDS speech samples achieved the highest correct identification rates, whereas their ADS speech samples were less well identified. In contrast, the majority of student-produced ADS speech samples were correctly identified by the raters, with IDS and FDS samples achieving lower correct identification rates (Table 3). The overall recognition

[5] More than 20% of the cells had expected frequencies of less than five. To deal with this problem we calculated the exact significance for each of the raters' chi-square results.

rate for actress-produced speech was 75%, whereas the recognition rate for student-produced speech was much lower (57.77%).

<insert Table 3>

Discussion

This study set out to evaluate the validity of standardised sentence production in comparative speech research. Our results suggest that successful reproduction of some of the acoustic characteristics of natural speech using simulated speech (e.g., speech elicited through descriptive scenarios and from speakers without interaction partner in the laboratory) may depend on the research paradigm ('simulated free speech' or standardised sentences) and also type of speaker (actresses or students).

Actresses

Consistent with our earlier findings in the 'simulated free speech' condition (Knoll et al., 2009), mean F_0 and vowel hyperarticulation measures produced by actresses in standardised sentences were similar to those found in natural interactions (Uther et al., 2007). If anything, it could be argued that vowel hyperarticulation of the actresses in the sentence condition revealed a marginal improvement compared to their vowel hyperarticulation in the 'simulated free speech' condition, in that it bears a greater resemblance to hyperarticulation values of natural speech. FDS had been positioned between IDS and ADS in the 'simulated free speech' condition[6] (Knoll et al., 2009)

[6] In that study, parametric tests showed that FDS vowel space was not significantly different from IDS or ADS vowel space, although non-parametric tests followed the trend of the natural speech samples, and showed a significant difference between FDS and ADS vowel space with no difference between IDS and FDS vowel space.

whereas in the present study FDS vowel space was significantly more expanded than ADS vowel space, but not IDS vowel space. Interestingly, actresses were also able to improve their performance for vowel duration in the sentence condition, with IDS exhibiting longer vowel durations than the two adult conditions, which is a close approximation to the results of the natural speech samples.

We propose the following reason for these marginal improvements in the sentence compared to the 'simulated free speech' condition. In our earlier study (Knoll et al., 2009), we suggested that actresses should be used to role play and improvisations, explaining their ability to produce some of these modifications in the 'simulated free speech condition', particularly compared to the student speakers. Nonetheless, most of the actresses' work is likely to be based on scripts, and the extra task of improvising their own script may have interfered with their ability to produce a wider range of modifications in the 'simulated free speech' condition. Providing them with standardised sentences might thus have released them to concentrate fully on the production of the relevant speech modifications rather than the production of a script. These findings add further support to the notion that actresses are capable of producing simulated speech that is similar to speech in natural interactions (e.g., Knoll et al., 2009; cf. Scherer 2003), and that their performance can even be improved by scripted speech.

Students

Interestingly, we were even able to approximate the mean F_0 findings previously found for natural speech (Uther et al., 2007) with standardised sentence production for our student speakers. This better approximation of mean F_0 is similar to previous research using simulated interactions (Biersack et al., 2005; Schaeffler et al., 2006), but is also an

improvement on the students' results for 'simulated free speech', which showed less similarity in F_0 to natural speech (Knoll et al., 2009). As in the case of the actresses, although to a greater degree, it is possible that the additional requirement for the participants to invent their own script in our previous study (Knoll et al., 2009) might have interfered with the student speakers' ability to convincingly reproduce acoustic characteristics normally found in natural speech.

Another aspect that might also explain the improvement in performance for mean F_0, is that the standardised sentences always followed each respective 'simulated free speech' condition. This was an intentional measure aimed to relax the participants and put them into the right mindset for each standardised sentence, similar to procedures used in basic emotional (cf. Scherer, 2003) and irony research (e.g., Rockwell, 2000). It is thus possible that a slight training effect may have taken place for the increased F_0. This would also be supported by the observation that the majority of student speakers found the sentence condition easier than the 'simulated free speech' condition. It is possible that this 'previous training' might have resulted in a more practiced and relaxed speech production, which consequently was perceived as less difficult by the speakers. However, this possibility remains to be confirmed.

Consistent with our findings for student 'simulated free speech' (Knoll et al., 2009), we were unable to replicate the occurrence of vowel hyperarticulation in both IDS and FDS in the standardised sentences condition. The occurrence of this hyperarticulation had been one of the main findings of Uther et al.'s (2007) natural speech study. It is surprising that the student hyperarticulation measures did not reflect the above discussed 'training effect', or benefit from the reduction of 'task load' to the same extent as the

student F_0 measures or the actresses' hyperarticulation measures. The reason for this could be explained by the differing levels of conspicuousness of these two acoustic characteristics. Increased F_0 (pitch) is a more distinctive feature of IDS that is easily perceived by a naïve listener, and might thus be more likely to be approximated in 'simulated interactions', particularly where a 'training effect' is present. In contrast, hyperarticulation has been suggested to be an unconscious speech modification (e.g., Burnham et al., 2002; Lindblom, 1990), which might be less discernable to a listener, and therefore more difficult to replicate in 'simulated interactions'. To date, the occurrence of vowel hyperarticulation seems to have only been reported in studies where the speaker was either *explicitly* instructed to produce clear speech (e.g., Smiljanić and Bradlow, 2005), in human-computer interaction error feedback tasks (e.g., Oviatt et al., 1998), with actresses (e.g., Knoll et al., 2009) or in comparative studies using real interaction partners who were perceived to be in need of some kind of linguistic modification (e.g., Burnham et al., 2002; Kuhl et al., 1997; Uther et al., 2007). It is likely that relevant 'error' feedback from a physically present interaction partner clearly perceived as needing clarity-enhancing linguistic modifications is crucial in eliciting this important feature in IDS (or any other relevant comparison speech recipient group), particularly if the speaker has no acting training or has not been specifically instructed to produce clear speech.

Previous exposure and ratings

Actresses and students did not report differing exposure to any of the speech recipient groups, suggesting that the differences in their performance were not due to their previous experience with any of the groups. Scherer (1986, p. 144) states that actresses might over-exaggerate obvious acoustic emotional cues, and miss other subtle ones.

Consistent with this view, we found that measures of actress vowel articulation and duration in the sentence condition were larger than those of the students and those found in the natural speech samples of Uther et al. (2007). This is also supported by the better rater identification of the different speech recipient groups for the actresses compared to the students. Most of the IDS and FDS speech samples for the actresses were correctly identified, whereas ADS achieved high confusion rates with IDS and FDS. It is possible that the actresses were exaggerating known stereotypical acoustic cues in those two speech types (e.g., high pitch in IDS, slow speech rate in FDS). In contrast, ADS as a speech type is less likely to have specific stereotypical acoustic cues, and this might have led to a poorer recognition rate for actress-produced ADS. It is also possible that the ADS speech samples were confused with IDS and FDS speech samples, because actress ADS speech samples comprised greater vowel articulation than *all* of the students speech samples. The recognition rates for student IDS and FDS were substantially lower than those of the actresses, but the recognition rate for ADS was relatively high for the student samples. This could be an artifact due to the majority of student speech samples being classified as ADS regardless of their actual speech recipient group, for instance, because of the smaller vowel articulation in the student speech samples compared to those of the actresses.

However, F_0 measures of the actresses were not significantly higher than those of the students and, if anything, their IDS F_0 measures were lower than those of the natural IDS speech samples (Uther et al., 2007). As previously mentioned, vowel hyperarticulation in speech is probably less conspicuous than F_0. Actresses' exaggeration of such a subtle acoustic cue (hyperarticulation) on the one hand, whilst not exaggerating

a more obvious cue (F_0) on the other hand, would seem to contradict the assumption (cf. Scherer, 2003) that actresses would fail to reproduce these more subtle cues.

Conclusions

Both student and actress speakers improved their reproduction of natural acoustic speech modifications using standardised sentences compared with their 'simulated free speech'. These improvements are most likely to relate to a reduced 'task load', although a slight training effect in the sentence (compared to the 'simulated free speech') condition is also possible. Further research is required to investigate the relative importance of these factors. The difficulty in replicating hyperarticulation in simulated conditions with untrained, inexperienced or uninstructed speakers, even with a reduced 'task load', raises the question of what makes this modification such a unique characteristic in terms of its occurrence in linguistically relevant speech. We suspect that a reciprocal dynamic feedback between speaker and listener may be an important factor in the process of generating appropriate speech modifications, particularly if the speaker lacks acting training. A trade-off between the realism of the speech material and the need to limit speech content variability due to experimental control is probably inevitable. However, our results do suggest that if the experimental set-up requires such control and an interaction partner is absent, actresses should be used in preference to students. Furthermore, scripted sentences should be used in preference to 'simulated free speech' scenarios, although the latter could be used as inductor dialogue.

Acknowledgements

We are grateful to two anonymous reviewers for helpful comments that greatly improved this manuscript. 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 research bursary from the Economic and Social Research Council (ESRC) to Monja Knoll.

References:

Anolli, L., Ciceri, R., Infantino, M.G., 2000. Irony as a game of implicitness: acoustic profiles of ironic communication. J. Psycholinguist. Res. 29(3), 275-311.

Anolli, L., Ciceri, R., Infantino, M.G., 2002. From "blame by praise" to "praise by blame". Analysis of vocal patterns in ironic communication. Int. J. Psychol. 37, 266-276.

Banse, R., Scherer, K.R., 1996. Acoustic profiles in vocal emotion expression. J. Pers. Soc. Psychol. 70, 614-636.

Bänziger, T., Scherer, K.R., 2005. The role of intonation in emotional expressions. Speech Commun. 46(3-4), 252-267.

Biersack, S., Kempe, V., 2005. Exploring the influence of vocal emotion expression on communicative effectiveness. Phonetica 62, 106-119.

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/.

Bradlow, A. R., Baker, R. E., Choi, A., Kim, M., and Van Engen, K. J. The Wildcat corpus of native- and foreign-accented English. J. Acoust. Soc. Amer. 121(5), 3072.

Bradlow, A.R., Kraus, N., Hayes, E., 2003. Speaking clearly for children with learning disabilities: sentence perception in noise. J. Speech, Lang. and Hear. Res. 46, 80-97.

Bradlow, A.R., Toretta, G.M., Pisoni, D.B., 1996. Intelligibility of normal speech I: global and fine-grained acoustic-phonetic talker characteristics. Speech Commun. 20(3-4), 255-272.

Breitenstein, C., Van Lancker, D., Daum, I., 2001. The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample. Cogn. Emot. 15(1), 57-59.

Burnham, D., Kitamura, C., Vollmer-Conna, U., 2002. What's new pussycat? On talking to babies and animals. Science 296, 1095.

Cheang, H.S., Pell, M.D., 2008. The sound of sarcasm. Speech Commun. 50, 366-381.

Cohen, A., Starkweather, J.A., 1961. Vocal cues to the identification of language. Am. J. Psychol. 74, 90-93.

Ferguson, S.H., Kewley-Port, D., 2002. Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. J. Acoust. Soc. Amer. 112(1), 259-271.

Friend, M, Farrer, M.J., 1994. A comparison of content-masking procedures for obtaining judgments of discrete affective states. J. Acoust. Soc. Amer. 1283-1290.

Jakobovits, L.A., 1965. Repetition of auditory presented information. Psychol. Rep. 17, 785-786.

Johnson, N. L., Kotz, S., 1970a. Continuous univariate distributions – 1. Chichester: John Wiley & Sons, Ltd.

Johnson, N. L., Kotz, S., 1970b. Continuous univariate distributions – 2. Chichester: John Wiley & Sons, Ltd.

Kalikow, D.N., Stevens, K.N., Elliott, L.L., 1977. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. J. Acoust. Soc. Amer. 61(5), 1337-1351.

Knoll, M.A, Scharrer, L., Costall, A., 2009. Are actresses better simulators than female students? The effects of simulation on prosodic modifications of infant- and foreigner-directed speech. Speech Commun. 51, 296-305.

Knower, F.H., 1941. Analysis of some experimental variations of simulated vocal expressions of the emotions. J. Soc. Psychol. 14, 369-372.

Kramer, E., 1964. Elimination of verbal cues in judgments of emotion from voice. J. Abnorm. Soc. Psychol. 68, 390-396.

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.

Ladd, D.R., Silverman, K.E.A., Tolkmitt, F., Bergmann, G., Scherer, K.R., 1985. Evidence for the independent function of intonation contour type, voice quality, and F_0 range in signalling speaker affect. J. Acoust. Soc. Amer. 78(2), 435-444.

Laukka, P., Juslin, P. N., Bresin, R., 2005. A dimensional approach to vocal expression of emotion. Cogn. Emot. 19(5), 633-653.

Lindblom, B., 1990. Explaining phonetic variation: A sketch of the H and H theory, in: W. Hardcastle and A. Marchal (Eds.), Speech Production and Speech Modelling, Kluwer Academic Publishers, Dordrecht, pp. 403–439.

Milmoe, S., Rosenthal, R., Blane, H.T., Chafetz, M.E., Wolf, I., 1967. The doctor's voice: postdictor of successful referral of alcoholic patients. J. Abnorm. Soc. Psychol. 72, 78-84.

Oviatt, S., Levow, D., MacEachern, M., Kuhn, K., 1998. Modeling hyperarticulated speech during human-computer error resolution. Speech Commun. 24(2), 87-110.

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. Appl. Psycholinguist. 12, 481–504.

Parkinson, B., 1996. Emotions are social. Br. J. Psychol. 87, 663-683.

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. J. Speech Hear. Res. 28, 96-103.

Rockwell, P., 2000. Lower, slower, louder: vocal cues of sarcasm. J. Psycholinguist. Res. 29(5), 483-495.

Rogers, P. L., Scherer, K.R., Rosenthal, R., 1971. Content filtering human speech: A simple electronic system. Behav. Res. Methods 3, 16-18.

Ross, M., Duffy, R.J., Cooker, H.S., Sargeant, R.L., 1973. Contribution of the lower audible frequencies to the recognition of emotions. Am. Ann. Deaf 118, 37-42.

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., 1971. Randomized splicing: a note on a simple technique for masking speech content. J. Exp. Res. 5, 155-159.

Scherer, K.R., 1986. Vocal affect expression: a review and a model for future research. Psychol. Bull. 99(2), 143-165.

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

Scherer, K.R., Banse, R., Wallbott, H.G., 2001. Emotion inferences from vocal expression correlate across languages and cultures. J. Cross. Cult. Psychol. 32, 76-92.

Scherer, K.R., Banse, R., Wallbott, H.G., Goldbeck, T., 1991. Vocal cues in emotion encoding and decoding. Motiv. Emot. 15, 123-148.

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

Scherer, K.R., Koivumaki, J., Rosenthal, R., 1972. Minimal cues in the vocal communication of affect: judging emotions from content-masked speech. J. Psycholinguist. Res. 1, 269-285.

Smiljanić, R., Bradlow, A.R., 2005. Production and perception of clear speech in Croatian and English. J. Acoust. Soc. Amer. 118 (3), 1677-1688.

Soskin, W.F., Kauffman, P.E., 1961. Judgment of emotion in word-free voice samples. J. Commun. 11, 73-80.

Starkweather, J.A., 1956. The communication-value of content-free speech. Am. J. Psychol. 69, 121-123.

Starkweather, J.A., 1967. Vocal behavior as an information channel of speaker status, in: Salzinger, K., Salzinger, S. (Eds.), Research in verbal behavior and some neurophysiological implications, Academic Press, New York, pp. 253-262.

Tischer, B., 1988. Ein Verfahren zur Messung der vokalen Kommunikation von Gefuehlen. Sprache Kogn. 7(4), 205-216.

Trainor, L.J., Austin, C.M., Desjardins, R.N., 2000. Is infant-directed speech prosody a result of the vocal expression of emotion? Psychol Sci. 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 Commun. 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. Percept. Mot. Skills 96, 759-771.

Wallbott, H.G., Scherer, K.R., 1986. Cues and channels in emotion recognition. J. Pers. Soc. Psychol. 51, 690-699.

Williams, C.E., Stevens, K.N., 1972. Emotions and speech: some acoustical correlates. J. Acoust. Soc. Amer. 52, 1238-1250.

Standardised sentence production is routinely used in speech research to avoid content variability typical of natural speech production. However, the validity of such standardised material is not well understood. Here, we evaluated the use of standardised sentences by comparing them to two existing, non-standardised datasets of simulated free and natural speech (the latter produced by mothers in real interactions). Standardised sentences and simulated free speech were produced by students and actresses without an interaction partner. Each dataset comprised recordings of infant- (IDS), foreigner- (FDS) and adult-directed (ADS) speech, which were analysed for mean F_0, vowel duration and hyperarticulation. Whilst students' mean F_0 pattern in standardised speech was closer to the natural speech than their previous 'simulated free speech', no difference in vowel hyperarticulation and duration patterns was found for students' standardised sentences between the three speech styles. Actresses' F_0, vowel duration and hyperarticulation patterns in standardised speech were similar to the natural speech, and a part improvement on their 'simulated free speech'. These results suggest that successful reproduction of some acoustic measures (e.g., F_0) can be achieved with standardised content regardless of the type of speaker, whereas the production of other acoustic measures (e.g., hyperarticulation) are context- and speaker-dependent.

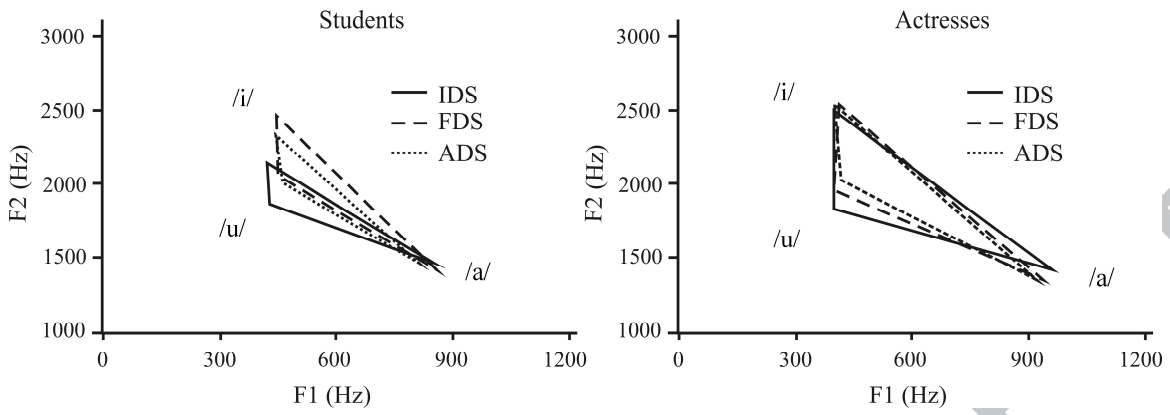


Fig1

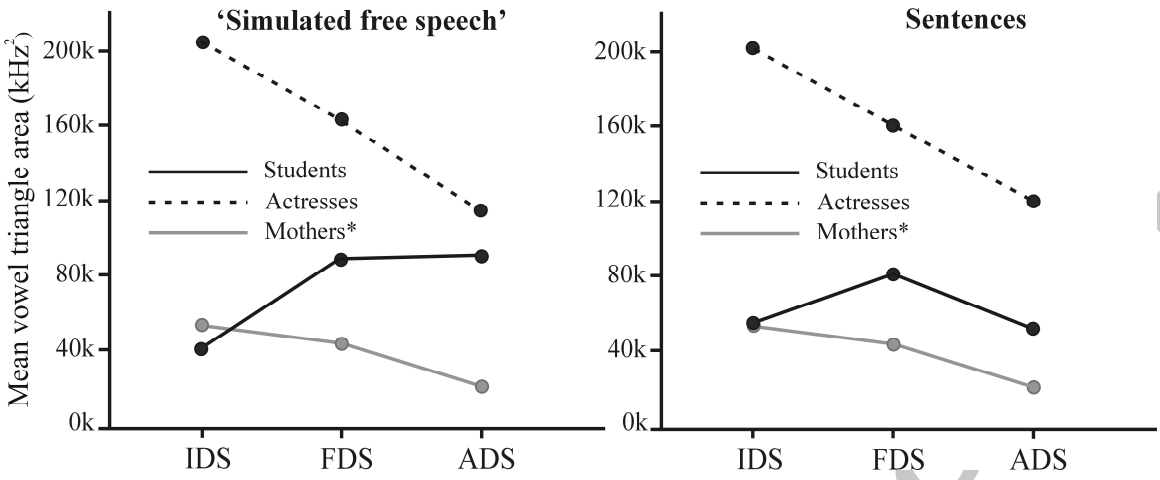


Fig2

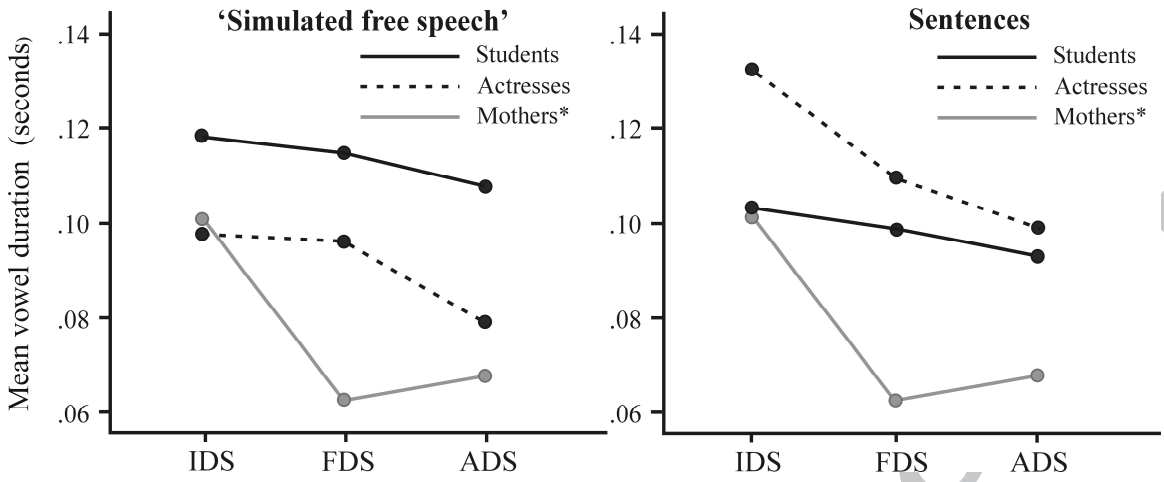


Fig3

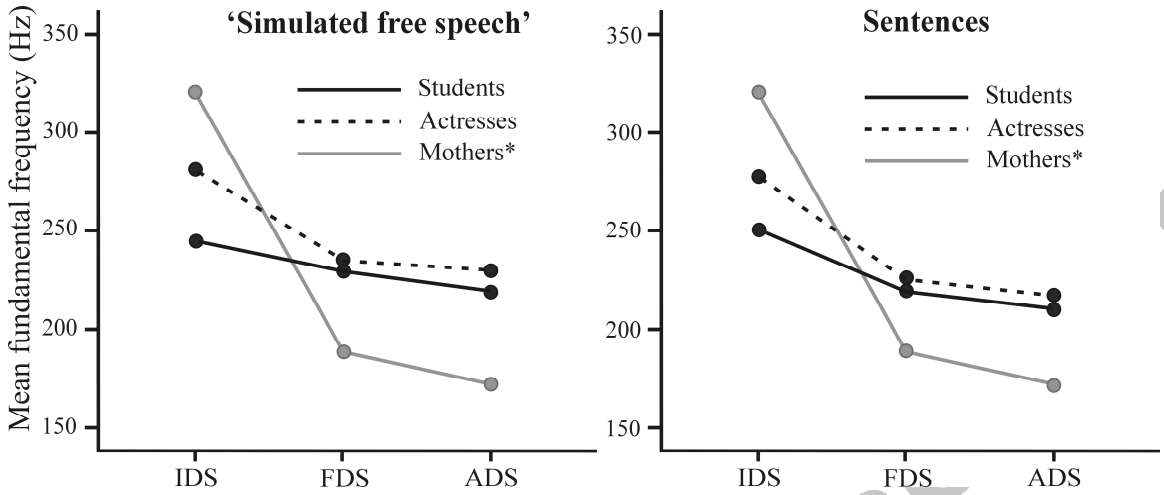


Fig4

Table captions:

Table 1

Comparison of findings of acoustic modifications in natural and simulated speech produced by actresses and students (after Knoll et al., 2009).

Table 2

Chi-square values, contingency co-efficient and exact significance for each of the raters' identification of IDS, FDS and ADS.

Table 3

Confusion matrix for type of speech recipient group (IDS, FDS and ADS) and rated speech recipient group for actresses and students (in percent). Shaded areas show percentages of correctly identified samples.

Table 1

	Vowel space	*Mean F_0*	*Vowel duration*
Natural speech samples *Uther et al. (2007)*	IDS & FDS greater than ADS. No difference between IDS & FDS.	IDS higher than ADS & FDS. No difference between ADS & FDS.	IDS longer than FDS & ADS. No difference between ADS & FDS.
Simulated speech produced by students (Knoll et al., 2009)	IDS smaller than FDS & ADS. No difference between FDS & ADS.	IDS greater than ADS. FDS positioned between IDS and ADS.	No difference between the three speech recipient groups.
Simulated speech produced by actresses (Knoll et al., 2009)	IDS greater than ADS. FDS positioned between IDS & ADS.	IDS greater than ADS & FDS. No difference between ADS & FDS.	No difference between the three speech recipient groups.

Unshaded cells represent aspects of simulated speech that are consistent with natural speech, light shaded areas indicate aspects that approximate natural speech, and darker shaded areas are inconsistent with natural speech.

Table 2

	Actresses			*Students*		
	χ^2	**Contingency co-efficient**	**Exact sig.**	χ^2	**Contingency co-efficient**	**Exact sig.**
Rater 1	22.681	.656	.001	9.521	.491	.045
Rater 2	27.982	.695	.001	15.5	.584	.002
Rater 3	36.154	.739	.001	12.75	.546	.01
Rater 4	29.167	.702	.001	6.5	.422	.171
Rater 5	14.238	.567	.005	10.621	.511	.027
Rater 6	30.539	.710	.001	10.621	.511	.027
Total	**154.993**			**66.68**		

Table 3

		Rated		
		IDS	**FDS**	**ADS**
Actresses	*IDS*	93.34	3.33	3.33
	FDS	8.33	83.34	8.33
	ADS	23.33	28.33	48.34
Students	*IDS*	48.34	28.33	23.33
	FDS	6.66	48.34	45.00
	ADS	6.66	18.34	75.00

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 students (left panel) and actresses (right panel).

Figure 2. IDS, FDS and ADS mean vowel area (calculated from F1/F2 vowel triangle space) for standardised sentence production (right) and 'free speech' production (left; Knoll et al., 2009) of students and actresses in comparison to mothers' natural speech production. *Values for mothers were adapted from Uther et al. (2007).

Figure 3. IDS, FDS and ADS mean vowel duration for standardised sentence production (right) and 'free speech' production (left; Knoll et al., 2009) of students and actresses in comparison to mothers' natural speech production. *Values for mothers were adapted from Uther et al. (2007).

Figure 4. IDS, FDS and ADS mean F_0 for standardised sentence production (right) and 'free speech' production (left; Knoll et al., 2009) of students and actresses in comparison to mothers' natural speech production. *Values for mothers were adapted from Uther et al. (2007).