Characterising $F_{(0)}$ contour shape in infant- and foreigner-directed speech

Monja A. Knoll, Alan Costall

Abstract

Previous research has used both natural and simulated interactions to investigate the functions of infant-directed speech (IDS), but the effect of these approaches on the results of such studies is unclear. The aim of this study was to compare $F_{(0)}$ contours of natural and simulated speech directed to three distinct speech recipient groups to investigate the effects of different methodologies on research findings. $F_{(0)}$ contours were derived from IDS, foreign-directed speech (FDS) and speech directed to a native adult speaker (ADS). The speech was produced by ten students and ten actresses in simulated 'interactions' with an imaginary speech partner and compared to an existing dataset of natural mother–infant interactions. The $F_{(0)}$ contours were analysed with a qualitative and a quantitative method. In the existing natural dataset, IDS $F_{(0)}$ contours (mainly sinusoidal–bell shapes) had been markedly different from FDS and ADS $F_{(0)}$ contours (mainly level shapes). However, these results were not reproducible in the simulated interactions with either students or actresses. Here, the different shapes were more evenly distributed across IDS, FDS and ADS with no difference between these three groups. Consequently, it is suggested that simulated speech should be used with caution in IDS or any other speech research.

1. Introduction

The majority of previous emotional or linguistic speech research has utilised simulated speech obtained in the laboratory, rather than genuine natural speech, to obtain standardised content that can be analysed and compared across different conditions (e.g., Banse and Scherer, 1996, Bänziger and Scherer, 2005, Biersack et al., 2005, Bradlow, 2002, Bradlow et al., 1996, Papousek and Hwang, 1991, Smiljanić and Bradlow, 2005 and Wallbott and Scherer, 1986). Different speaker groups have also been used to obtain these simulated speech samples. Students have mostly been used because they are an easily available 'source' of participants in psychological research, and they represent a generic type of participant (e.g., Biersack et al., 2005). Another group, actresses, has been used because they are thought to be better than lay-people at portraying emotional speech (e.g., Banse and Scherer, 1996, Scherer, 1986, Scherer, 2003 and Wallbott and Scherer, 1986).

Some researchers have recognised and discussed the potential problems associated with the controlled nature of such simulated speech in comparison with natural speech (e.g., Murray and Arnott, 1993), and the desirability of using spontaneous (natural) speech has also been highlighted by Smiljanić and Bradlow (2005). However, direct comparisons of the simulated speech of students and actresses with natural speech have rarely been carried out (but see Williams and Stevens, 1972). Recently, we attempted to address this

lack of knowledge when 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 those produced by actresses and students in two previous studies (Knoll et al., 2009 and Knoll et al., 2011). The current study is the third in a line of studies that aim to investigate comparability between simulated and natural genuine interactions focusing on $F_{(0)}$ contours occurring in IDS, ADS and FDS.

1.1. IDS background

Previous research has established that adults talk differently in interactions with infants compared to conversations with other adults (e.g., Burnham et al., 2002, Fernald, 1993, Fernald and Kuhl, 1987 and Uther et al., 2007). The acoustic and phonetic characteristics of IDS are well recognised and considered to be universal (e.g., Fernald, 1989, Grieser and Kuhl, 1988, Kuhl, 2000 and Kuhl et al., 1997). IDS has a characteristically higher fundamental frequency (F_0) than ADS, and contains exaggerated F_0 contours, shorter utterances, longer pauses and hyperarticulation of vowels (e.g., Andruski and Kuhl, 1996, Bernstein Ratner, 1986, Bernstein Ratner and Luberooff, 1984, Burnham et al., 2002, Fernald and Simon, 1984, Karzon, 1985, Kuhl et al., 1997, Stern et al., 1983, Trainor et al., 2000 and Uther et al., 2007).

Acoustic modifications in IDS probably have a linguistic role, but they might also have emotional or attention-gaining functions (e.g., Bernstein Ratner, 1986, Fernald and Simon, 1984, Fernald, 1993, Grieser and Kuhl, 1988, Karzon, 1985, Kitamura and Burnham, 2003, Kuhl, 2000 and Trainor et al., 2000).

Proponents of the *linguistic* role suggest that native acoustic language input during early development alters the neural processing of the incoming speech signal resulting in the creation of perceptual categories for native phonemes (i.e. mental maps; see Kuhl, 2000 and Kuhl et al., 1997 for a full review). This perceptual mapping of incoming speech biases infants' phonetic perception towards the sounds of their native language (e.g., Best, 2002, Best et al., 1995, Kuhl et al., 1997 and Tsao et al., 2006). Due to this, infants at 6 months fail to discriminate foreign language contrasts that they were able to discriminate earlier, and language input thus clearly has a major effect on infants' perception of phonemic categories. For instance, research (e.g., Burnham et al., 2002 and Kuhl et al., 1997) has shown that mothers use hyperarticulation¹ of vowels in IDS, which seems to emphasise the difference between the phonemes of their native language. This may serve to make phonetic mapping more effective (e.g., Kuhl et al., 1997 and Kuhl, 2000), and also appears to aid infants' ability to discriminate phonemes (Lui et al., 2003).

Because infants do not understand the lexical meaning of words, advocates of the affective role suggest that the exaggerated features of IDS communicate emotional content to infants (e.g., Trainor et al., 2000). In this model the higher mean F_0 and exaggerated F_0 contours in IDS convey positive affect to the infant, and facilitate social interaction (e.g., Fernald, 1993, Singh et al., 2002 and Trainor and Desjardins, 2002). IDS

might also play an *attentional* role where the exaggerated acoustic features of IDS serve to engage and maintain infant attention, and also encourage social turn taking (e.g., Fernald and Kuhl, 1987 and Fernald and Mazzie, 1991).

The functions of IDS have been investigated by comparing IDS with unemotional and emotional ADS (e.g., Fernald and Simon, 1984, Fernald and Mazzie, 1991, Kuhl et al., 1997 and Trainor et al., 2000), emotional pet-directed speech (PDS; Burnham et al., 2002) and FDS as an unemotional linguistic comparison group (e.g., Biersack et al., 2005, Levitt and Geffrion, 1994, Papousek and Hwang, 1991, Knoll et al., 2007 and Uther et al., 2007).

Focusing specifically on the linguistic role of $F_{(0)}$ contours, Knoll et al. (2007) investigated $F_{(0)}$ contour variations in infant- and foreigner-directed speech. The authors analysed natural interactions between ten mothers and their infants, and a foreign and native British confederate. In line with previous research, they found that IDS was characterised by exaggerated $F_{(0)}$ contours, whereas ADS and FDS were characterised by level contours. Knoll et al. (2007) concluded that the function of the exaggerated F_0 contours was probably emotional or attentional. In the present study we will use Knoll et al. (2007) as the comparative natural data-set.

1.2. Impact of different methodologies on research findings

Different methodologies have been used to obtain speech samples in comparative IDS research, such as natural interactions in the mothers' homes (e.g., Burnham et al., 2002, Englund and Behne, 2005 and Uther et al., 2007), and simulated interactions in the laboratory with mothers (e.g., Trainor et al., 2000) and non-mothers or students (e.g., Papousek and Hwang, 1991 and Biersack et al., 2005). The extent to which these different methodologies and the use of simulated speech in comparison to genuine natural speech have affected the outcome of these studies is not entirely clear. Certainly, research on the facial expression of emotion gives good reason to investigate differences between *posed* and spontaneous speech. For instance, formants 1 and 2 production can be mediated by facial musculature, and therefore facial expression is important in generating acoustic features (e.g., those associated with emotion). Furthermore, research into the facial expression of emotion has long distinguished between voluntary and involuntary expressions (see Prodger, 2009, p. 16, on Darwin and Duchenne). There are also differences in form and timing between *posed* and *spontaneous* expressions that can be distinguished by observers (Davis and Gibson, 2000, Ekman and Friesen, 1982, Hess and Kleck, 1990, McLellan et al., 2010 and Parkinson, 1995). Furthermore, genuine expressions of emotions activate established neural pathways that allow rapid and more complete emotional expression in the face and by extension in the voice. It therefore follows that there are definite neuroanatomical differences (Rinn, 1984 and Matsumoto and Lee, 1993) involved in

generation of facial expression during production of speech, and it is possible that certain acoustic features that are related to these in genuine natural speech may be difficult or impossible to replicate in posed (i.e. simulated) speech because of these neural differences.

Another problem with this area of research is that comparison studies are relatively rare. 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 spontaneous speech produced by mothers), but that simulated IDS presented more exaggerated F_0 contours than ADS. In contrast, Schaeffler et al. (2006) compared acoustic modifications in speech of mothers in interactions with their children ('spontaneous speech') and non-mothers without an interaction partner ('non-spontaneous speech'). These authors found that spontaneous speech differed in terms of shimmer and harmonics-to-noise ratio from *non-spontaneous* speech, but not in relationship to any other acoustic measures (i.e., F_0 , formant 1, formant 2, etc.). However, Schaeffler et al. (2006) used standardised brochures and sentences in both their conditions including the 'spontaneous' condition, rendering the spontaneity and naturalness of that condition doubtful. Consequently, more comparative investigations of both natural and simulated based speech research are clearly needed.

We recently attempted to determine whether different methods of data collection resulted in comparable speech modifications by comparing target words of natural speech samples of IDS, ADS and FDS with those produced by students and actresses in imaginary interactions using 'free speech' (without a script; Knoll et al., 2009) and standardised sentences (Knoll et al., 2011). We found that we could reproduce the rated high emotional affective aspects of the different natural speech types in the simulated laboratory condition (IDS > ADS > FDS; established by affective ratings of low-pass filtered excerpts of the actress and student speech samples) in both studies. We speculated at the time that the speakers in imaginary IDS may still have used exaggerated F_0 contours, and that this may have been used as a cue by the raters for the identification of the higher emotional affect in simulated IDS. Furthermore, the speech produced by actresses exhibited vowel hyperarticulation and mean F_0 measures similar to the natural speech samples in both studies. However, we were strikingly unsuccessful in reproducing vowel hyperarticulation in our student speech samples, although the students were able to simulate the characteristically high mean F_0 in IDS in the standardised sentence condition (Knoll et al., 2011). Vowel hyperarticulation has been found to be a linguistic characteristic of both IDS and FDS (e.g., Uther et al., 2007), but the modification may not be easily perceived and reproduced by speakers without specific experience (e.g., acting training) or direct interaction with the target group (e.g., mothers with infants). The students, due to lack of acting training, were possibly diffident or unable to produce hyperarticulation of vowels due to the subtlety of the modification, and would be more successful with the reproduction of such a conspicuous cue as exaggerated $F_{(0)}$ contours. We intend to investigate this possibility in the present study.

1.3. Aim of the research

The main aim of this study is to further investigate the degree of comparability between simulated and natural speech by exploring whether $F_{(0)}$ contour patterns occurring in natural IDS, FDS and ADS can be reproduced in simulated interactions with student and actress speakers. In previous studies we already analysed and compared other acoustic measures across the natural and simulated datasets (see Knoll et al., 2009 and Knoll et al., 2011). For this reason and to keep our study comparable to the published findings of Knoll et al. (2007), we focused our analyses on $F_{(0)}$ contours in the present study. As actresses had already been found to be able to reproduce similar acoustic features in simulated IDS and FDS to natural speech in our earlier studies, we would expect patterns of F_0 contour variation between IDS, FDS and ADS produced by actresses to closely resemble those of the mothers (i.e., exaggerated contours in IDS, with level contours in FDS and ADS). The ability of students to replicate natural $F_{(0)}$ contours in simulated speech is presently unknown. However, in contrast to the trained actresses, we would expect that the untrained students' may be less able to replicate accurately these F_0 contours.

In order to facilitate optimal comparison between the natural and the two simulated datasets, the current study utilised the student and actresses' 'free speech' (e.g. Knoll et al., 2009) rather than the 'standardised sentences' datasets (Knoll et al., 2011) for the extraction of the $F_{(0)}$ contours. In the sentence condition, the speakers were instructed to utter the sentences 'Look at the shark' (or sheep or shoe). Such sentences have an imperative discursive function with the target word appearing at the final position of the utterance. By comparison, genuine IDS, ADS and FDS utterances will include a wide variety of different discursive functions and target word positions in those utterances. We hypothesised that the use of the 'free speech' condition would provide the actresses and students with the freedom to produce a variety of functional utterances that are comparable with those of the mothers' natural speech. We tested this by analysing the discursive function of the utterance and the position in which the target word appeared across the speaker and speech recipient groups.

2. Method

The acoustic measures of the present study were derived from two existing datasets of speech directed to infants, foreigners, and adults. The two datasets consisted of simulated speech samples produced in 'imaginary interactions' by students (henceforth referred to as 'students') and by professional actresses (henceforth referred to as 'actresses'; Knoll et al., 2009 and Knoll et al., 2011). These two data-sets were compared to an existing data-set of F_0 contours that were derived from natural interactions (henceforth referred to as 'mothers'; Knoll et al., 2006 and Knoll et al., 2007). Specific information about the speakers can be found in the original papers, but the essential details are presented here.

2.1. Speakers

2.1.1. Natural dataset – ‘mothers’ (Knoll et al., 2007)

The natural speech samples consisted of recordings of ten mothers from Southern England (mean age 30.7 years) interacting with their infants, a foreign (Chinese) and southern English confederate (both women in their early twenties). The recordings were carried out in the mothers' houses. For each interaction, the mothers were provided with three toys to promote consistency of conversation content, otherwise the interactions were natural and spontaneous.

2.1.2. Simulated speech – ‘students’ and ‘actresses’

The simulated speech samples consisted of recordings of ten female students (mean age 22.9 years) and ten professional actresses (mean age 28.4 years), who imagined talking to an infant, a foreign and a British adult. The actresses had all undergone professional acting training and had been acting for at least 1 year (mean 5.6 years). The students reported no previous acting experience. None of the speakers was hearing impaired, and they had been living in the South of England for at least 80% of their lives. To keep these imaginary situations consistent with the speech samples of the mothers, the speakers were instructed to imagine that they were talking to a close family member (e.g., niece, nephew or even own child) in the infant condition, but they were not provided with an example or idea of how IDS should sound. For both 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 had been living in the UK for less than 2 months, and that they might encounter some communication problems. In order to avoid sources of acoustic variation due to different utterances (a problem in non-standardised speech samples), the speakers were provided with the same toys as the mothers in the original study to promote use of the same target words ‘sheep’, ‘shark’ and ‘shoe’ in each condition. Apart from the toy material, the speakers were encouraged to construct and invent their own scenarios, which ideally should have resulted in ‘free speech’.

2.1.3. Raters

Ten raters (five for each of the speaker groups) were recruited from the student and staff population of the University of Portsmouth for the qualitative analysis. We recruited different raters for each of the speaker groups (students and actresses), as the analyses for these were not carried out simultaneously. In order to determine reliability, Cohen's Kappa for each of the rater pairs² was calculated, and the overall median and mean values for these calculations for each speaker group were determined. The inter-rater reliability

between the other raters was found to be sufficient (see Table 1 for gender distribution, age, mean and median Cohen's Kappa). To determine the intra-rater reliability, one rater in each of the speaker groups repeated the procedure 3 weeks later, and this was found to be good (students = .748; actresses = .870).

Table 1.

Gender-distribution, mean age (*sd*) and median and mean values of Cohen's Kappa for each of the raters in each speaker condition.

	Gender distribution	Mean age (<i>sd</i>)	Median and mean Cohen's Kappa
Students	Three females	31.2 (5.9)	.610 (mean)
	Two males		.575 (median)
Actresses ^a	Three females	29.0 (7.1)	.670 (mean)
	One male		.661 (median)

a Ratings of rater 4 were removed after this process, as this rater's results did not show good agreement with the other raters (mean Cohen's Kappa before removal: .416). Qualitative analyses for the actresses are therefore based on four raters.

2.2. Extraction of $F_{(0)}$ contours and image files

To retain consistency with Knoll et al. (2007), our $F_{(0)}$ contour analyses focused on specific target words ('shark', 'sheep', 'shoe'). The use of target words was necessary to allow comparison across the different speech recipient (IDS, FDS and ADS) and speaker conditions (mothers, students and actresses).

It should be noted that, although contour analysis of whole utterances might be informative in an overall sense, even studies that have analysed phrases have focused detailed information extraction at the single word level (Fernald and Simon, 1984 and Stern et al., 1983). Linguistic (acoustic) modifications at the word level have already been demonstrated in IDS (e.g., Andruski and Kuhl, 1996, Burnham et al., 2002 and Kuhl et al., 1997), FDS (e.g., Uther et al., 2007), and even ADS (e.g., Fernald and Maizzie, 1991). It has also been shown that infants (e.g., Bull et al., 1984, Eilers et al., 1984 and Nazzi et al., 1998) and adults (e.g., Bradlow et al., 1996) are sensitive to and can benefit from word-level acoustic changes, supporting the notion that further word-level linguistic highlighting may be present in the form of specific $F_{(0)}$ contours shapes.

From an original dataset of 763 words, 360 of the target words (i.e. *sheep*, *shark*, *shoe*) were randomly selected for the extraction of the $F_{(0)}$ contours. This reduction in data was necessary because of the time-intensive nature of the image-processing phase and also made it comparable to the methods of Knoll et al. (2007). This resulted in an assignment of 180 words per speaker condition (students and actresses), divided into 60 words per speech recipient condition (IDS, ADS and FDS). $F_{(0)}$ contours were extracted from these words using Praat 4.5.16 (Boersma and Weenink, 2006), the $F_{(0)}$ range for the extraction was set at a floor of 100 Hz and a ceiling of 600 Hz. In order to obtain more homogeneous contours, the 'smooth' function

was used (set at bandwidth 10 Hz). Finally, the $F_{(0)}$ contours were graphically rendered in Praat 4.5.16 (set at a duration of 0.7 s). $F_{(0)}$ contours were then standardised for duration without distorting their shape, and the lines were then converted to a standard TIFF format at 72 dpi in Corel PhotoPaint 11.0. Our methods here follow Knoll et al. (2007).

2.3. Rating scheme

The rating scheme was adapted from Knoll et al. (2007) and comprised five shapes consisting of (1) bell (up-down); (2) complex (sinusoidal up-down-up or down-up-down); (3) falling (down); (4) rising (up); and (5) level shapes. The raters were also offered an 'undecided' option that they could select if they were unable to categorise the images into one of the five shapes.

2.4. Procedure

2.4.1. Qualitative analysis

For the qualitative analysis, raters were required to allocate each of the presented $F_{(0)}$ contour images to one of the shapes on the rating scheme. Raters were told that the shapes on the rating scheme were 'ideal' representations and that the presented images would probably not exactly resemble these 'ideal' shapes, and that they could select the 'undecided' option if unsure. Before the main rating, raters were given trial images (which were not included in the later analysis) to familiarise themselves with the images and rating scheme. Once this trial was completed, the raters were presented with the images in random order to avoid order effects.

2.4.2. Eigenshape Analysis

The use of Eigenshape Analysis for $F_{(0)}$ contour analysis along with qualitative analysis was first introduced and evaluated by Knoll et al., 2006 and Knoll et al., 2007 and the relevant background can be found in these papers. However, as this technique is normally used in the zoological sciences for biological morphology categorisation rather than acoustic research, a short introduction is in order. Eigenshape Analysis uses a principal component-like approach to characterise the actual shape variation of ($F_{(0)}$) contours through transformation of the contours' x-y co-ordinates (MacLeod, 1999). TpsDig 2.0 (Rohlf, 2005) was used to collect these x-y contour coordinate nodes from the $F_{(0)}$ contour TIFF images. Digitisation begins from a common landmark point located at the start of a shape, and this arrangement must remain constant for all images to be digitised (MacLeod, 1999). Consistent with the direction of Cartesian graphing output, in the current study the extreme left hand side was used as the starting point, and the nodes were positioned

along the upper edge of the $F_{(0)}$ contour. The number of coordinate points required to characterise the curve accurately depends on the complexity of its shape, and must remain constant in all images. To keep our methods consistent with Knoll et al. (2007) we used 37 coordinate points, which were found to be sufficient to characterise all curves.

The resulting coordinate pairs were then transformed from their Cartesian (x–y) form to a Φ (Phi) shape function using MacLeod’s ‘X–Y to Phi.exe’ program. This procedure provides a series of net angular deviations from the starting point that represent a dimensionless map of the contour shape (MacLeod, 1999). Based on an accuracy level of 95% (see MacLeod, 1999), nine nodes (IDS = nine; ADS = seven; FDS = eight) for the students and eight nodes (IDS = eight; ADS = seven; FDS = five) for the actresses were interpolated from the original 37-node curves. This is in contrast to 18 nodes (IDS = 16; ADS = three; FDS = three) in the natural data-set (Knoll et al., 2007). This indicates that, in general, the shapes in the natural dataset were more complex than the shapes in the student and actress datasets. The Φ -transformed coordinate data output of ‘X–Y to Phi.exe’ was then analysed using ‘Eigenshape.exe’. We provided a summary of the different steps related to Eigenshape Analysis and the respective Eigenshape programs in Table 2, but see Knoll et al. (2007) for specific information on this procedure.

Table 2.

Summary of the function and resulting output of the Eigenshape Analysis programmes (adapted from Knoll et al., 2007).

Programme	Function	Output
‘TPS-dig’	Collection of the contour coordinate nodes from the pitch contour TIFF images	Output of x–y co-ordinate pairs
‘X–Y to Phi.exe’	Transformation of co-ordinate pairs from their Cartesian (x–y) form to a Φ (Phi) shape function	Output consists of transformed co-ordinate data
‘Eigenshape.exe’	Analysis of transformed co-ordinate data	Output includes: the mean Eigenshape, Eigenvalues and the individual Eigenshape scores on each Eigenshape axis (output can be used for discriminant analysis)
‘Phi to X–Y.exe’	Graphic visualisation of each Eigenshape and the mean Eigenshape	Output can be used for graphic visualisation of shape

2.4.3. Analysis of discursive function and word position

In addition to the qualitative and quantitative analysis of the F_0 contours, coding was also carried out of the discursive functions of the utterances in which the analysed words appeared, together with word position within these functions. As no comparative data existed, the coding was carried out for each of the three datasets. The coding scheme for the discursive function included four statement types, which were

declarative statements, interrogative statements, imperative statements and exclamations. This was sufficient to code the functions in the present study. Four different word positions were also included in this coding scheme consisting of words at the beginning of an utterance including accompanying articles (utterance-initial), words in the middle of an utterance (utterance-medial), words at the end of an utterance (utterance-final) and stand-alone utterances of the target words.

2.5. Results

For the qualitative analysis, the ratings of the $F_{(0)}$ contours (frequencies) were subjected to Pearson's Chi-square analysis. The 'undecided' category was rarely chosen, and was therefore omitted from the analysis. This resulted in a 3 (IDS, FDS and ADS) $\times$ 5 (contours) contingency table. Eigenshape scores generated from x-y coordinates in Eigenshape Analysis were analysed using CVA.

2.5.1. Qualitative analysis

A similar pattern of association existed across the different raters within each speaker condition. Consequently, each rater's rating for each image was taken into account, resulting in five data points per image per condition for the students and four data points per image per condition for the actresses (e.g. 60 IDS images in the student condition $\times$ 5 raters = 300 data points). Separate Chi-square analyses for the actress and student data were conducted. Knoll et al. (2007) reported a significant association between the contour category and speech recipient group for the natural data-set ($\chi^2 = 544.038$, $df = 8$, $p < .001$). This was also the case for student ($\chi^2 = 67.452$, $df = 8$, $p < .001$) and for actress speakers ($\chi^2 = 54.062$, $df = 8$, $p < .001$). However, greater discrimination³ between variables was found for the mothers (contingency coefficient: .628) than for the students (contingency coefficient: .264; $F(8, 8) = 8.066$, $p < 0.01$) or for the actresses (contingency coefficient: .267; $F(8, 8) = 10.063$, $p < 0.01$). However, there was no difference between the actresses and the students. To provide a further comparison between the different contingency tables and to account for varying sizes, we calculated Cramer's V. Cramer's V produced a value of .196 for the actresses and .160 for the student speakers, compared to the reported value of .571 for the mothers in the natural data-set (Knoll et al., 2007).

Similar to the results for the mothers, a higher proportion of IDS produced by actresses and students was characterised as bell and complex contours (27.6% and 28% respectively) than ADS or FDS (see Table 3). However, the difference between IDS and the adult conditions was significantly smaller than in the natural speech samples (88%). A large proportion of the IDS contours for students and actresses were also characterised as level contours (31% and 27.2% respectively) compared to a smaller proportion of the mothers' IDS contours (16%). Consistent with the mothers' data, the actresses and students produced

fewer bell or sinusoidal contours in both FDS (students: 25.1%, actresses: 24.5%, mothers: 1.5%) and ADS (students: 14%, actresses: 19.8%, mothers: 1.1%) compared to IDS. A large proportion of ADS contours produced by students were characterised as level contours (39.8%) followed by falling contours (26.1%) and then rising contours (20.1%). This proportion is still not comparable to the natural speech samples, where the frequency of level contour ratings was much higher in the mothers' ADS (81.1%).

Table 3.

Distribution of ratings for each of the four contours across IDS, FDS and ADS for students, actresses and mothers. Mother data was adapted from Knoll et al. (2007).

Speech recipient	Contour types														
	Bell			Complex			Falling			Rising			Level		
	O	E	%	O	E	%	O	E	%	O	E	%	O	E	%
<i>IDS</i>															
Mothers	191	71.1	63.7	36	12.9	12.0	10	31.0	3.3	14	15	4.7	48	167.7	16.0
Students	40	23.5	13.8	69	50.3	23.8	45	67.9	15.5	46	57.8	15.9	90	90.5	31.0
Actresses	57	31.9	24.6	31	31.5	13.4	65	86.8	28.0	16	19.1	6.9	63	62.9	27.2
<i>FDS</i>															
Mothers	4	64	1.5	0	11.6	0	36	28.1	13.3	17	12	6.3	224	154.2	78.9
Students	26	24.2	8.7	49	51.9	16.4	85	70	28.4	71	59.6	23.7	68	96.3	22.7
Actresses	16	32.6	6.8	42	32.2	17.7	90	88.6	38.0	29	19.5	12.2	60	64.1	25.3
<i>ADS</i>															
Mothers	3	62.8	1.1	0	11.4	0	47	27.6	15.5	6	11.7	2.3	231	151.4	81.1
Students	6	24.2	2.0	36	51.9	12.0	78	70	26.1	60	59.6	20.1	119	93.3	39.8
Actresses	24	32.6	10.1	23	32.2	9.7	109	88.6	46.0	13	19.5	5.5	68	64.1	28.7

Note. O = Observed, E = Expected, % refer to percentages within speech recipient groups.

The majority of ADS $F_{(0)}$ contours produced by the actresses were characterised as falling contours (46%), followed by 28.7% level contours with a comparatively low percentage of rising contours (5.5%). In contrast to the FDS contours produced by mothers (mainly level contours), the majority of FDS contours for the actresses and students were evenly distributed between the categories (with the lowest percentage occurring in the bell category for both speaker groups, see Table 3). Our qualitative results here indicate that the $F_{(0)}$ contour distribution between the three speech recipient groups are markedly different in the natural dataset compared to the simulated datasets.

2.5.2. Eigenshape Analysis

2.5.2.1. Results provided by the Eigenshape programs

Almost 54% of the shape variation in all three speech recipient groups (IDS, FDS and ADS) for the student interactions, and 69% for the actress interactions, was attributable to the first Eigenshape axis (this information is provided by the program and is a measure of the shape variation between speech recipient conditions). The actress data here are comparable to that of the mothers, where almost 70% was attributable to the first Eigenshape axis. Separate analysis of IDS, ADS and FDS demonstrated that the variation on the remaining Eigenshape axes was due to all speech recipient conditions for both students and actresses. In contrast, the variation of the remaining Eigenshape axes in the mothers' dataset was entirely due to IDS, whereas ADS and FDS $F_{(0)}$ contour variation was more similar to each other (Knoll et al., 2007). This indicates that the three speech recipient groups for both students and actresses are similar in that they possess less distinct shape variation than those of the mothers.

Consistent with the qualitative analysis, the mean Eigenshape of the mothers' IDS $F_{(0)}$ contours had a more exaggerated curve than either ADS or FDS (see Fig. 1a), whereas the mean shapes of ADS and FDS exhibited almost identical slightly curved level contours. In contrast to this, the mean shapes of the three groups for the student interactions are relatively similar to each other and form inverted bell curves (Fig. 1b).

Fig. 1.

Comparison of mean pitch contour shapes produced by mothers (A), students (B) and actresses (C) for IDS, ADS and FDS plotted on Eigenshape 1 versus Eigenshape 2.

Complementary to the qualitative analysis, these mean shapes are due to the high occurrence of level, rising and falling contours in each of the three groups, which would result in this average outline. It is interesting to note that the adult $F_{(0)}$ contours of the actresses resemble those of the adult $F_{(0)}$ contours of the students, but to a certain degree the IDS $F_{(0)}$ contours resemble the natural IDS $F_{(0)}$ contours (Fig. 1a–c). The ADS and FDS contours of the simulated speech samples (students and actresses) were qualitatively characterised with more bell and complex contours than the ADS and FDS contours of natural speech, and these results are also observable in the mean shapes of the students and actresses (see Fig. 1a–c).

2.5.2.2. Canonical Variates Analysis (CVA)

To determine how distinct the three speech recipient groups are from each other, a CVA (simultaneous entry; separately for student and actress speakers) was performed with speech recipient groups as the independent variable, and the scores on the seven axes⁴ (students) and six axes (actresses) derived from Eigenshape Analysis as the dependent variables. Two discriminant functions were calculated for each of the two speaker groups (student and actresses), but these functions were not statistically significant for the students or for the actresses.

In contrast, Knoll et al. (2007) reported that the values of the first of these functions for the mothers were significantly different for IDS, ADS and FDS, whereas the second function was not significantly different. Fig. 2 illustrates how each of the $F_{(0)}$ contours cluster on functions one and two, and shows the group centroids for each speech recipient condition separately for mothers (adapted from Knoll et al., 2007, Fig. 2a), students (Fig. 2b) and actresses (Fig. 2c). Compared to the natural speech samples where the position of the group centroids indicated that IDS was very distinct from both ADS and FDS, the student and actress datasets show no distinction between the three speech recipient conditions.

Fig. 2.

Combined plot of CVA functions for IDS, ADS and FDS of mothers (A), students (B), and actresses (C).
Figure options

Overall, the discriminant functions predicted a successful classification for 44.4% of all student cases and for 45% of all actress cases, compared to almost 75% in the natural speech samples, see Table 4). Correct classification for each speech recipient group for the students and actresses was much lower than for the mothers (IDS: 75%, FDS: 59% and ADS: 77%). Interestingly, there was an even distribution between the three speech recipient conditions for students and for actresses with regards to which group was mistaken for which. For instance 30% of student IDS shapes were mistaken for ADS shapes, and 23.3% were mistaken for FDS shapes (30.0% and 21.7% respectively for the actress IDS shapes). The distribution in the adult conditions was similar to this, although the actress FDS shapes were more often confused with ADS than with IDS. The quantitative results provided by the Eigenshape programs and the subsequent CVA show that

there was better separation between IDS, FDS and ADS in the natural dataset compared to the simulated datasets.

Table 4.

Predictions of discriminant functions for each of the three speech groups (IDS, ADS and FDS) for students, actresses and mothers in percentage.

Speech groups		Predicted speech groups		
		IDS	FDS	ADS
IDS	Students	46.7	23.3	30.0
	Actresses	48.3	21.7	30.0
	Mothers	75.0	13.3	11.7
FDS	Students	26.7	46.7	26.7
	Actresses	15	56.7	28.3
	Mothers	5.6	77.0	25.9
ADS	Students	28.3	26.7	45.0
	Actresses	33.3	36.7	30.0
	Mothers	1.9	32.7	59.3

Italicised values refer to percentage agreement between the actual and predicted speech groups.

2.5.3. Comparison of qualitative and quantitative analyses

To determine the comparability between the raters' F_0 contour categorisations and the results from the Eigenshape programs, we conducted three CVA's (simultaneous entry, separate for the mothers, students and actresses). We determined the average shape for each of the contours based on the majority of the raters' categorisations and used these as the independent variable with the scores on the sixteen axes (mothers), seven axes (students) and six axes (actresses) derived from Eigenshape Analysis as the dependent variables. Four discriminant functions were calculated for each of the three speaker groups, two of these were significant for the mothers (Function 1: Chi-square = 305.981, $p < .001$, Wilks Lamda = .140; Function 2: Chi-square = 89.843, $p < .001$, Wilks Lamda = .561), three were significant for the students (Function 1: Chi-square = 158.087, $p < .001$, Wilks Lamda = .401; Function 2: Chi-square = 64.175, $p < .001$, Wilks Lamda = .690; Function 3: Chi-square = 28.236, $p = .002$, Wilks Lamda = .849) and two for the actresses (Function 1: Chi-square = 95.506, $p < .001$, Wilks Lamda = .573; Function 2: Chi-square = 45.923, $p < .001$, Wilks Lamda = .767). Fig. 3 shows the position of the group centroids for each of the contours for each of the speaker groups. The group centroids in the three datasets show good separation between each of the contours indicating that the scores on the Eigenshape program correspond with the categorisations of the raters. However, categorisation was slightly better in the natural compared to the simulated datasets. Overall, 91.6% of the cases were correctly identified in the natural compared to 67.9% in the

student and 62.8% in the actress datasets (see Table 5 for classification results). These results indicate that there was good to very good agreement between the raters' categorisation on the one hand and the results provided by the Eigenshape programs on the other.

Fig. 3.

Combined plot of CVA functions for qualitative contour types as categorised by the raters of mothers (A), students (B), and actresses (C).

Table 5.

Predictions of discriminant functions for contour types as categorised by the raters for students, actresses and mothers in percentage.

Speech groups	Contour types	Predicted contour types				
		Bell	Complex	Rising	Falling	Level
Mothers	Bell	97.6	0.0	0.0	2.4	0.0
	Complex	0.0	85.7	0.0	0.0	14.3
	Falling	0.0	0.0	<i>100.0</i>	0.0	0.0
	Rising	11.1	0.0	0.0	38.9	50.0
	Level	1.1	0.0	0.0	0.0	98.9
Students	Bell	53.8	7.7	0.0	23.1	15.4
	Complex	0.0	50.0	6.7	40.0	3.3
	Falling	0.0	7.0	55.8	4.7	32.6
	Rising	2.8	11.1	0.0	69.4	16.7
	Level	1.7	1.7	6.9	5.2	84.5
Actresses	Bell	60.0	6.7	20.0	13.3	0.0
	Complex	15.0	45.0	20.0	15.0	5.0
	Falling	1.4	2.8	69.0	1.4	25.4
	Rising	13.3	0.0	40.0	33.3	13.3
	Level	0.0	4.5	20.5	2.3	72.7

Italicised values refer to percentage agreement between the actual and predicted contour types.

2.5.4. Discursive function and word position

In order to investigate discursive function of utterance and word position as a function of speaker groups, speech recipient groups and F_0 contours types, we conducted twelve frequency analyses (i.e., Fisher's Exact Test, because assumptions of Pearson's Chi-square were not met).

2.5.4.1. Association between F_0 contour shape, function of utterance and word position

Given that the speech was spontaneous/unscripted in all of the conditions, it is possible that there were also linguistic differences – word position and discursive function – and that these were responsible, at least in part, for the observed acoustic differences. In fact, this was not the case. There was no significant association between F_0 contour shape (i.e., bell, complex, falling, rising and level contours) and word position (i.e., utterance-initial, utterance-medial, utterance-final and standalone) for any of the three speaker datasets. This means that word positions were equally distributed across the F_0 contour shapes for the three speaker groups.

There was a significant association between the contour shape and the function of the utterance for the mothers ($\chi^2 = 47.370$, $df = 6$, $p < .001$), the students ($\chi^2 = 21.546$, $df = 6$, $p < .007$) and the actresses ($\chi^2 = 18.903$, $df = 6$, $p < .024$). In the natural dataset, there were fewer bell and complex contours observed in declarative sentences than expected, but more than expected were observed for interrogative utterances (bell and complex contours) and exclamations (bell contours only). More rising and level contours were observed in declarative sentences than expected, with fewer than expected in interrogative utterances and exclamations (level contours only). In the student dataset, more falling contours were observed than expected in the declarative utterances, but fewer complex contours were observed in those utterances. However, more complex contours than expected were observed in interrogative and exclamation utterances. For the actress dataset, fewer bell and complex contours were observed than expected in declarative sentences, with more rising and level contours than expected. More bell contours were observed than expected in the interrogative utterances. In fact, there are similarities in the distribution of the expected and observed frequencies across the three speaker sets. Most importantly, the majority of the words used for the extraction of the $F_{(0)}$ contours occurred in declarative utterances. This was the case for all three speaker groups (mothers: 79%, students: 85% and actresses: 82.8%). However, in the natural dataset only 56% bell contours occurred in declarative utterances compared to 76.9% in the student and 66.7% in the actress datasets. The overall percentage distribution across the other discursive functions is comparable for the three speaker sets.

2.5.4.2. Speech recipient groups/function of utterance/word position

We also investigated whether there was an association between the speech recipient groups (i.e., IDS, ADS and FDS) and the discursive functions and word positions. In the natural dataset, there was a significant association between the speech recipient groups and the type of functional utterance ($\chi^2 = 40.124$, $df = 6$, $p < .001$) and the word positions ($\chi^2 = 18.909$, $df = 6$, $p < .001$). In IDS, mothers used fewer declarative utterances and more exclamations and interrogative and imperative utterances than expected. This was reversed for both adult conditions. Mothers also used fewer utterance-medial positions and more utterance-final positions than expected in the IDS condition, whereas the reverse was true for the adult conditions. More utterance-initial positions than expected were used in FDS compared to ADS where fewer than expected were used. There was a significant association between the speech recipient groups and the type of functional utterance ($\chi^2 = 19.123$, $df = 6$, $p < .001$) and the word positions ($\chi^2 = 12.719$, $df = 6$, $p < .032$) for student, but not for actress speakers. Students used fewer declarative utterances in IDS than expected and more interrogative utterances and exclamations. This was reversed for the adult conditions. This result is similar to the results of the natural dataset. Interestingly, students used more imperative utterances than expected in FDS, but fewer than expected in ADS. Students also used more median-utterance positions than expected for IDS and FDS, with fewer than expected in ADS. They also used more utterance-initial positions than expected in FDS and more utterance-final positions than expected in ADS.

The results for the discursive function and word position across the F_0 contour shapes and the speech recipient groups show that linguistic differences were not responsible for the observed acoustic differences between the natural and simulated datasets.

2.6. Discussion

The aim of the present study was to investigate the comparability of natural and simulated speech by comparing $F_{(0)}$ contours in IDS, FDS and ADS across three different speaker groups (mothers, students and actresses). In the present study, the difference in $F_{(0)}$ contours between the mothers' natural IDS (mainly sinusoidal–bell contours) and ADS/FDS (mainly level contours) were not reproducible in a laboratory context with simulated speech produced by either students or actresses. This finding was evident in the Eigenshape Analysis, as no difference was observed between the three speech recipient groups for both actresses and students, and in the results of the qualitative analysis, which markedly differed from those of the mothers' natural dataset (Knoll et al., 2007). Our finding that both methods are consistent with each other provides even stronger evidence of the difference between the simulated and natural datasets. Furthermore, our analyses of the function of the utterances and the word position of the target words indicate that our $F_{(0)}$ contour results were not an artefact of linguistic differences.

With regards to the students, the results are consistent with our earlier findings of vowel hyperarticulation in the same sample, but do not follow our findings with regards to rated emotional affect (Knoll et al., 2009). As with the natural data-set, we had previously found that student-produced IDS was rated more positively than FDS or ADS. We speculated that this may be due to a conspicuous cue (e.g., exaggerated $F_{(0)}$ contours). This was not the case, as a significantly smaller percentage of the overall student $F_{(0)}$ contours were characterised as bell and complex contours in the qualitative analysis compared to the natural samples and the same applied to the actresses. Trainor et al. (2000) noted an association between these shapes and particular emotional and turn-taking interactions, and suggested that these contour shapes were responsible for the melodic quality of IDS. This could explain why both speaker groups were unable to replicate these contours successfully, as they lacked an infant interaction partner to take turns with.

Whilst the result of student speakers was not entirely unexpected, given our previous findings (Knoll et al., 2009 and Knoll et al., 2011), we were surprised that the actresses also failed to simulate the pattern distribution of naturally occurring IDS and FDS $F_{(0)}$ contours. The failure of the actresses to reproduce $F_{(0)}$ contours similar to those of the mothers is particularly remarkable, as we had previously found that actresses were able to simulate most of the acoustic features occurring in IDS and FDS in both the current 'free speech' (Knoll et al., 2009) and in a 'standardised sentences' condition (Knoll et al., 2011). The finding is also puzzling considering our expectation that actresses might be able to tap into involuntary neural pathways for emotional expression, which should result in more natural displays of emotional expression in the face and thus voice. These results show that it may be impossible to simulate the full range of prosodic modifications in simulated speech even with professional acting training.

Interestingly, rather than simulated IDS revealing similar level contours to natural ADS and FDS, there was more variation within the simulated speech recipient conditions for both student- and actress-produced speech, and a wider distribution across the shapes than across those of the natural sample. A large proportion of ADS and FDS shapes used by the students and actresses seem to occupy the falling contour category, which is associated with declarative sentences in adult-directed speech (Stern et al., 1983). We found that the student and actress speakers mostly uttered the target words in such a manner (100% of falling contours for students; 80% for actresses). Because they provided their imaginary adult counterpart with a declarative statement (i.e., "This is a shark") or a description of the toy (e.g., "The shark has big teeth"). Interestingly, this was concordant with the mothers' $F_{(0)}$ contour pattern, as the second highest category for the natural ADS and FDS was falling contours. The students also used a high percentage of rising contours when addressing their imaginary foreign counterpart. Rising contours are associated with questioning and turn taking (Fernald, 1989), and it is possible that the students used this questioning speech style for the target words in their imaginary interactions as verification of comprehension. We

found that the students used more rising contours than expected in interrogative utterances, and this would support this suggestion.

It should be noted that the mothers used more rising contours in the FDS condition than in both the ADS and IDS conditions, which could lend support to the notion that rising contours may be characteristic of FDS. A higher degree of questioning might appear to cast aspersions on the listener's capacity to understand, and this finding may thus explain the higher degree of negative vocal affect for FDS in our earlier study (Knoll et al., 2009). Interestingly, we did not find that mothers used more imperative statements than expected in the FDS conditions. However, this may be due to the use of basic target words in an adult condition, and this possibility would need further investigation with fewer basic and more diverse target words. It has been found that once words are established as old information, they will no longer be linguistically highlighted (Fernald and Mazzie, 1991) and this may have occurred in the Knoll et al. (2007) FDS dataset.

Our findings of simulated speech also differ from those of previous studies investigating $F_{(0)}$ range in simulated IDS, FDS and ADS (Biersack et al., 2005), where IDS had been found to contain a greater $F_{(0)}$ range than the adult conditions. Although $F_{(0)}$ range and $F_{(0)}$ contours are associated, they are not synonymous, and the two studies are thus not directly comparable (e.g. a bell shaped $F_{(0)}$ contour could present the same $F_{(0)}$ range as a rising $F_{(0)}$ contour). However, it is possible that the $F_{(0)}$ range rather than the contour shape in our simulated conditions (students and actresses) would show a closer similarity to the $F_{(0)}$ range of the mothers, and we are currently in the process of investigating this possibility. Nevertheless, the $F_{(0)}$ contour shape of the simulated speech should have resembled that of the natural speech samples if the two are truly comparable. That this is not the case suggests that simulated speech of either actresses or students is not informative about the respective functions of $F_{(0)}$ contours in natural IDS and FDS.

Imaginary scenarios and simulated interactions are therefore not ideal for eliciting the full range of prosodic modifications found in natural speech, and in some cases even acting experience is insufficient to inform the speaker about the full modifications required in interactions with different audiences. It has already been noted that prosodic modifications in IDS are reflexive, instinctive and perhaps even unconscious (e.g., Burnham et al., 2002 and Kitamura and Burnham, 2003). If so, experience with infants may not strictly be necessary so long as a real infant is present to interact with. We also used toys for eliciting the target words in the present study to obtain the same target words as in the comparative natural speech study. It has been argued that toys can provide a focus of common attention and activity between parents and their children, and thus may have a crucial role in developing social skills and communication (e.g., Rodríguez, 2007 and Williams et al., 1999). Our findings show that the use of toys for

eliciting target words containing acoustically relevant modifications is only advisable in a context in which the speaker can use the toy as an interactive medium to encourage joint attention and communication.

Similarly, a foreign confederate with genuine comprehension problems and more adult-appropriate tasks that are relevant within an interaction between a foreign and native speaker may be required to elicit spontaneous speech with all the clarity-enhancing speech modifications. After all, a real speech partner will provide genuine comprehension feedback as part of normal natural speech, and the speaker can then react to this feedback with the appropriate speech modification regardless of whether it is intended to be linguistic, emotional or attentional. We mentioned earlier that Murray and Arnott (1993) have highlighted the difficulties of the 'trade off' between the 'realism' of speech and the necessity of controlled content in speech research, and that the desirability of using spontaneous speech in research has also been acknowledged by Smiljanic and Bradlow (2005). Recently, there have been attempts at providing tasks that elicit spontaneous speech whilst retaining some degree of control in the form of target words (e.g., Diapix task; Van Engen et al., 2010, and the DiapixUk task, which is an extension of the original Diapix task; Baker and Hazan, 2011). The use of these tasks has already been successfully implemented in forensic speech research (e.g. Knoll and Thorburn, 2012) and other studies comparing clear reading and clear speech styles (Hazan and Baker, 2010). The present study highlights the need for the development of more similarly innovative tasks, and for more speech research to utilise the Diapix task.

2.7. Conclusions

This was the third of a series of studies investigating the comparability of natural and simulated speech using IDS, ADS and FDS as the focus of investigation. In contrast to our previous studies (Knoll et al., 2009 and Knoll et al., 2011), neither the students nor the actresses in the present study were able to approximate the $F_{(0)}$ contour distribution found in natural IDS (sinusoidal–bell contours) and the two adult conditions (mainly level contours). Therefore simulated speech interactions should be used with caution in future research. Reciprocal feedback between speakers and listeners seems to be necessary to elicit the full range of speech modifications, and there is clearly a need for the development of more ecologically valid speech scenarios to investigate spontaneous speech.

Acknowledgements

This research was supported by a studentship grant from the ESRC – United Kingdom to Monja Knoll. We thank Lauren Gillies (UWS, UK) for help with the analyses and Stig Walsh (NMS, UK) for help with technical support and discussion. We also thank two anonymous reviewers for helpful comments on an earlier version of the manuscript.

References

- Andruski, J.E., Kuhl, P.E., 1996. The acoustic structure of vowels in mothers' speech to infants and adults. In: Proceedings of the Fourth International Conference on Spoken Language Processing, Philadelphia, PA, pp. 1545–1548.
- R. Baker, V. Hazan DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs Behav. Res. Methods, 43 (3) (2011), pp. 761–770
- R. Banse, K.R. Scherer Acoustic profiles in vocal emotion expression J. Pers. Soc. Psychol., 70 (1996), pp. 614–636
- T. Bänziger, K. Scherer The role of intonation in emotional expressions Speech Commun., 46 (3–4) (2005), pp. 252–267
- N. Bernstein Ratner Durational cues which mark clause boundaries in mother–child speech J. Phonet., 14 (1986), pp. 303–309
- N. Bernstein Ratner, A. Luberoff Cues to post-vocalic voicing in mother–child speech J. Phonet., 12 (1984), pp. 285–289
- C.T. Best Revealing the mother tongue's nurturing effects on the infant ear Infant Behav. Dev., 25 (2002), pp. 134–139
- C.T. Best, G.W. McRoberts, R. LaFleur, J. Silver-Isenstadt Divergent developmental patterns for infants' perception of two nonnative consonant contrasts Infant Behav. Dev., 18 (1995), pp. 339–350
- 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, Lisbon, pp. 2401–2405.
- Boersma, P., Weenink, D., 2006. Praat: Doing Phonetics by Computer (Version 4.5.16) <<http://www.praat.org/>>.
- A. Bradlow Confluent talker- and listener-related forces in clear speech production Lab. Phonol., 7 (2002), pp. 241–273

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

D. Bull, R.E. Eilers, D.K. Oller Infant's discrimination of final syllable fundamental frequency in multisyllabic stimuli *J. Acoust. Soc. Am.*, 77 (1984), pp. 289–295

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

R.E. Eilers, D.H. Bull, D.K. Oller, D.C. Lewis The discrimination of vowel duration by infants *J. Acoust. Soc. Am.*, 75 (1984), pp. 1213–1218

P. Ekman, W.F. Friesen Felt, false and miserable smiles *J. Nonverbal Behav.*, 6 (4) (1982), pp. 238–252

K.T. Englund, D.M. Behne Infant directed speech in natural interaction – Norwegian vowel quantity and quality *J. Psycholinguist. Res.*, 34 (2005), pp. 259–280

A. Fernald Intonation and communicative intent in mothers' speech to infants: is the melody the message? *Child Dev.*, 60 (1989), pp. 1497–1510

A. Fernald Approval and disapproval: infant responsiveness to vocal affect in familiar and unfamiliar languages *Child Dev.*, 64 (1993), pp. 657–674

A. Fernald, P. Kuhl Acoustic determinants of infant preference for motherese speech *Infant Behav. Dev.*, 10 (1987), pp. 279–293

A. Fernald, C. Mazzie Prosody and focus in speech to infants and adults *Dev. Psychol.*, 27 (2) (1991), pp. 209–221

A. Fernald, T. Simon Expanded intonation contours in mother's speech to newborns *Dev. Psychol.*, 20 (1984), pp. 104–113

P.J. Davis, M.G. Gibson Recognition of posed and genuine facial expressions of emotion in paranoid and nonparanoid schizophrenia *J. Abnorm. Psychol.*, 109 (3) (2000), p. 445

D.L. Grieser, P. Kuhl Maternal speech to infants in a tonal language: support for universal prosodic features in motherese *Dev. Psychol.*, 24 (1) (1988), pp. 14–20

Hazan, V., Baker, R., 2010. Does reading clearly produce the same acoustic-phonetic modifications as spontaneous speech in a clear speaking style? In: *Proceedings of DiSS-LPSS, Tokyo, Japan*, pp. 7–10.

U. Hess, R. Kleck Differentiating emotion elicited and deliberate emotional facial expressions *Eur. J. Soc. Psychol.*, 5 (1990), pp. 369–385

R.G. Karzon Discrimination of polysyllabic sequences by one to four-month-old infants *J. Exp. Child Psychol.*, 39 (1985), pp. 326–342

C. Kitamura, D. Burnham Pitch and communicative intent in mother's speech: adjustments for age and sex in the first year *Infancy*, 4 (1) (2003), pp. 85–110

Knoll, M.A., Thorburn, E., 2012. My voice lies: The use of acoustic measures in lie detection. In: *Paper Presented at the Annual Conference of the BPS Cognitive Section in Glasgow*.

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

M.A. Knoll, L. Scharrer, Costall "Look at the shark": evaluation of student- and actress-produced standardised sentences of infant- and foreigner-directed speech *Speech Commun.*, 53 (2011), pp. 12–22

Knoll, M.A., Uther, M., MacLeod, N., O'Neill, M., Walsh, S., 2006. Emotional, linguistic or cute? The function of pitch contours in infant- and foreigner-directed speech. In: *Proceedings of the 3rd International Conference on Speech Prosody, Dresden, Germany*, pp. 165–168.

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

P.K. Kuhl Language, mind, and brain: experience alters perception M.S. Gazzaniga (Ed.), *The New Cognitive Neuroscience* (second ed.), MIT Press, Cambridge (2000), pp. 99–115

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

A.G. Levitt, J. Geffrion Prosodic characteristics of register variation in the read speech of adult native speakers of French (A) *J. Acoust. Soc. Am.*, 96 (5) (1994), pp. 3349–3350

H.M. Liu, P.K. Kuhl, F.M. Tsao An association between mothers' speech clarity and infants' speech discrimination skills *Dev. Sci.*, 6 (2003), pp. F1–F10

N.M. MacLeod Generalizing and extending the eigenshape method of shape space visualisation and analysis *Paleobiology*, 25 (1) (1999), pp. 107–138

L.A. Marascuilo Statistical Methods for Behavioural Science Research McGraw-Hill Book Company, New York (1971)

D. Matsumoto, M. Lee Consciousness, volition, and the neuropsychology of facial expressions of emotion *Conscious. Cogn.*, 2 (3) (1993), pp. 237–254

T. McLellan, L. Johnston, J. Dalrymple-Alford, R. Porter Sensitivity to genuine versus posed emotion specified in facial displays *Cogn. Emot.*, 24 (8) (2010), pp. 1277–1292

I.R. Murray, J.L. Arnott Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion *J. Acoust. Soc. Am.*, 93 (2) (1993), pp. 1097–1108

T. Nazzi, C. Floccia, J. Bertoncini Discrimination of pitch contours by neonates *Infant Behav. Dev.*, 21 (4) (1998), pp. 779–784

M. Papousek, S.-F.C. Hwang Tone and intonation in Mandarin babytalk to presyllabic infants: comparison with registers of adult conversation and foreign language instruction *Appl. Psycholinguist.*, 12 (1991), pp. 481–504

B. Parkinson Ideas and Realities of Emotion Psychology Press (1995)

P. Prodger Darwin's Camera: Art and Photography in the Theory of Evaluation Oxford University Press (2009)

W.E. Rinn The neuropsychology of facial expression: a review of the neurological and psychological mechanisms for producing facial expressions *Psychol. Bull.*, 95 (1) (1984), p. 52

C. Rodríguez Object use, communication, and signs: the triadic basis of early cognitive development
J. Valsiner, A. Rosa (Eds.), *The Cambridge Handbook of Sociocultural Psychology*, Cambridge University Press, Cambridge (2007), pp. 257–276

Rohlf, F.J., 2005. TpsDig, Digitise Landmarks and Outlines (Version 2.0) [Computer Program]. Department of Ecology and Evolution, State University of New York at Stony Brook <<http://life.bio.sunysb.edu/morph/>>.

Schaeffler, F., Kempe, V., Biersack, S., 2006. Comparing vocal parameters in spontaneous and posed child-directed speech. In: *Proceedings of the 3rd International conference of Speech Prosody*, Dresden, Germany, pp. 688–691.

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

K.R. Scherer Vocal communication of emotion: A review of research paradigms *Speech Commun.*, 40 (2003), pp. 227–256

L. Singh, J.L. Morgan, C.T. Best Infants' listening preferences: baby talk or happy talk? *Infancy*, 3 (2002), pp. 365–394

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

D.N. Stern, S. Spieker, R.K. Barnett, K. MacKain The prosody of maternal speech: infant age and context related changes *J. Child Lang.*, 10 (1983), pp. 1–15

L.J. Trainor, C.M. Austin, R.N. Desjardins Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychol. Sci.*, 11 (3) (2000), pp. 188–195

L.J. Trainor, R.N. Desjardins Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels *Psychon. Bull. Rev.*, 9 (2) (2002), pp. 335–340

F.-M. Tsao, H.-M. Liu, P.K. Kuhl Perception of native and non-native affricate–fricative contrasts: cross-language tests on adults and infants *J. Acoust. Soc. Am.*, 120 (4) (2006), pp. 2285–2294

M. Uther, M.A. Knoll, D. Burnham Do you speak E-n-g-l-i-s-h? A comparison of foreigner- and infant-directed speech *Speech Commun.*, 49 (1) (2007), pp. 2–7

K. Van Engen, M. Baese-Berk, R.E. Baker, A. Choi, M. Kim, A.R. Bradlow The wildcat corpus of native- and foreign-accented English: communicative efficiency across conversational dyads with varying language alignment profiles *Lang. Speech*, 53 (4) (2010), pp. 510–540

H.G. Wallbott, K.R. Scherer Cues and channels in emotion recognition *J. Pers. Soc. Psychol.*, 51 (1986), pp. 690–699

C.E. Williams, K.N. Stevens Emotions and speech: some acoustical correlates *J. Acoust. Soc. Am.*, 52 (1972), pp. 1238–1250

E. Williams, A. Costall, V. Reddy Children with autism experience problems with both objects and people *J. Autism Dev. Disord.*, 29 (5) (1999), pp. 367–378

Part of this work is included in the Proceedings of the 4th International Conference on Speech Prosody (pp. 256–259). Campinas, Brazil.

1 Vowel hyperarticulation as used here, refers to the expansion of the two-dimensional space between the corner vowels /a/, /i/ and /u/ due to articulatory formant 1 and formant 2 changes (see Kuhl et al., 1997).

2 Cohen’s Kappa was calculated between all the raters (i.e. ratings of rater 1 were compared with rater 2, then rater 3, etc. resulting in ten Cohen Kappa scores for each speaker condition). We then calculated the mean and median of these ten scores for the three speaker conditions separately.

3 To compare whether the strength of the relationship between the variables differed significantly between the three speech groups, we calculated the F-value from the ratio of their respective Chi-square/degrees of freedom values (see Marascuilo, 1971, pp. 330–331 for the relationship between the Chi-square and F-distribution).

4 Although, the program calculated two more axes for each of the speaker conditions, these last two axes consisted of a high proportion of zero values and are thus not included in the analysis (MacLeod, 1999).