

Emotional, linguistic or just cute? The function of pitch contours in infant- and foreigner-directed speech

¹Monja Knoll, ¹Maria Uther, ²Norman MacLeod, ²Mark O'Neill & ²Stig Walsh

¹SAIL, Department of Psychology, University of Portsmouth, Portsmouth, UK

²Department of Palaeontology, The Natural History Museum, London, UK

monja.knoll@port.ac.uk

Abstract

This study evaluated the relative functions of pitch contours in infant-directed speech (IDS) by comparing it with adult-directed (ADS) and foreigner-directed speech (FDS). The shape of pitch contours derived from target words in speech samples was analysed using two novel algorithmic methods and a standard qualitative approach. Our findings indicate that IDS is very distinct from ADS and FDS, whilst the latter two exhibit a strong similarity to each other. These results suggest that pitch contours in IDS serve an emotional-attentional function rather than a linguistic function.

1. Introduction

Infant-directed speech (IDS) is characterised by the presence of hyperarticulation, increased pitch, exaggerated pitch contours, shorter utterances, longer pauses and high emotional affect [1-4]. These characteristics are well recognised and seem to be universal [5]. It is generally believed that IDS probably serves at least three different functions, but no consensus has been reached on their relative importance or independence. These three roles can be broadly divided into an attentional, an emotional and a linguistic function [5]. Previous research has investigated these roles by comparing IDS with non-emotional adult-directed speech (ADS) mainly in imaginary interactions with the help of scenarios [2]. However, ADS as a comparison to IDS is not sufficient as ADS lacks both the emotional and linguistic requirements of IDS. Recent research has therefore compared IDS with both emotional pet-directed speech (PDS) [4] and foreigner-directed speech representing a linguistic comparison group [6]. Foreign-directed speech might present a particularly valid linguistic comparison group as speakers might modify their speech due to perceived comprehension difficulties (language and accent). Using these methods it was found that hyperarticulation (expansion of F1/F2 space) of vowels has a linguistic role [6] and that increased pitch has an emotional function [4].

One area that was not explored in those studies was the nature and function of the exaggerated pitch contours in IDS compared to FDS. This was partially addressed in earlier studies with reference to tonal [7] and non-tonal languages [8]. Pitch range (an aspect of the pitch contour) was found to increase in FDS but not in IDS in the tonal language (Mandarin Chinese), whereas the opposite was apparent in the non-tonal language (English). Because pitch is used to convey linguistic meaning in tonal languages, these findings suggest that pitch contours might have a linguistic function. However, both of these studies investigated the pitch range and relied on imaginary scenarios, and it remains unclear to what extent pitch contour *shape* in those conditions actually differs in genuine natural interactions.

Pitch contours have routinely been analysed qualitatively using human raters [2]. There have been several attempts to find reliable alternative quantitative methods to pitch contour analysis [e.g. 3, 9, 10], but no single approach has emerged as either simple in its execution, or capable of being applied to a wide variety of speech types. Approaches using mathematical modelling of pitch values [9] allow analysis of a variety of different speech components, but in practice proved rather complex for general use. However, recent experimental work based on principle component analysis of syllable pitch values (eigenpitch [10]) avoids much of this complexity. Although the method remains to be tested in comparative speech conditions, it offers the potential to characterise pitch contours in a generalised context. Careful scaling of contours would be required to avoid inclusion of other non-shape related variables (e.g. contour length, referential pitch) without over abstraction of the contour shape, and as such the approach is presently not suitable for accurate characterisation of pitch contour shape.

Here we evaluate two approaches that are currently used in the natural sciences and that might provide objective easily usable alternative methods of pitch contour analysis. The first is Eigenshape Analysis (EA), which also uses a principal component-like approach to characterise the actual shape variation of pitch contours in series of conditions through transformation of the contours' x-y co-ordinates [11]. The second approach comprises an artificial neural net system that is presently in development as an automated identification system at the Natural History Museum (NHM), London. The Digital Automated Identification SYstem (DAISY) works by grouping objects it perceives to share similar features, allowing the shape-space consistency of pre-defined groupings (e.g. IDS, ADS and FDS) to be tested using the system's own diagnostic functions [12].

The present study has two aims. The first is to determine whether exaggerated pitch contours in IDS have a linguistic or emotional function by comparing them with a linguistic FDS comparison group (with ADS as a control group). The second aim is to evaluate the utility of the two algorithmic approaches for pitch contour analysis by comparing them to the more traditional approach of using qualitative analysis.

2. Method

2.1. Data collection and preparation of image files

Speech samples from ten British mothers recorded during natural interactions with their infants, a foreign and British confederate (both adults) [6] were analysed. A procedure was used to ensure that speakers used the target words 'sheep', 'shark' and 'shoe' in each condition. These words had been found to be prosodically highlighted in earlier studies [4, 6], and pitch contours were extracted from these. A total of 167

words were used for the extraction of the pitch contours (IDS = 60, ADS = 53, FDS = 54). Pitch contours were extracted in Praat 4.1.19 [13], the pitch range for the extraction was set at a floor of 100 Hz and a ceiling of 600 Hz (recommended setting in Praat 4.1.19 for female voices) [13]. In order to obtain more homogenous contours, the ‘smooth’ function was used (set at bandwidth 10 Hz). Finally, the pitch contours were drawn in Praat 4.1.19 (set at a pitch range of 100-600Hz, duration of 0.7 seconds).

Pitch contours were then standardised for duration without distorting the shape, the lines were enhanced to 8-point thickness, and then converted to a standard 500 x 500 pixel grid TIFF format at 72 dpi in Corel PhotoPaint 11.0. The reason for this procedure was to create a standardised format that could be used in all three approaches. The line thickness enhancement was required for the DAISY analysis, as the system is only effective using images with strong patterns of pixel contrast.

2.2. Procedure

2.2.1. Qualitative analysis

Five raters (four females, one male; mean age 29.6, SD 5.4) were used to rate and categorise the 167 pitch contour images using a rating scheme with five shapes and one undecided option. The shapes consisted of 1) bell; 2) complex; 3) falling; 4) rising, and 5) level shapes (see additional material), chosen on the basis of previous studies [2]. Raters were informed that the shapes on the rating scheme were ideal representations, and that the presented images would probably not exactly resemble these ‘ideal shapes’. Raters were given five trial images to familiarise themselves with the images and rating scheme. Once this trial was completed, the raters were presented with the 167 images in counterbalanced order to avoid order effects. The order was reversed for the last two raters to avoid anomalous results for the final images. The raters were not aware that those shapes constituted three different speech conditions. No time limit was given for the procedure. In order to determine reliability, intraclass correlation was carried out for the five raters (reliability coefficient $\alpha = 0.97$), and was found to be good. To determine the intra-rater reliability, rater number two repeated the procedure three weeks later, and this was also found to be good (reliability coefficient $\alpha = 0.96$).

2.2.2. Eigenshape analysis

TpsDig 2.0 [14] was used to collect the contour coordinate nodes from the pitch contour TIFF images. The number of coordinate points required to accurately characterise the curve depends on the complexity of its shape, must remain constant in all images and begins at a common landmark point (detailed discussion of the eigenshape technique can be found elsewhere [11]). In the current study 37 coordinate points were found to sufficiently characterise all curves, and the extreme left hand side was used as the starting point. The coordinate pairs were then transformed from their Cartesian (x-y) form to a Φ shape function using MacLeod’s ‘X-Y to Phi.exe’ program to provide a series of net angular deviations from the starting point that represents a dimensionless map of the contour shape [11]. Based on the recommended accuracy level of 95% [11], 18 nodes were interpolated from the original 37 node curves. The Φ -transformed coordinate data output of ‘X-Y to Phi.exe’ was then analysed using ‘Eigenshape.exe’. The output of this program includes determination of each eigenshape, the mean eigenshape, eigenvalues and the individual eigenshape scores on each eigenshape axis (used for further statistical analysis).

X-y coordinates that can be plotted for graphic visualisation of each eigenshape and the mean eigenshape were provided by ‘Phi to X-Y.exe’. Eigenscores for each eigenshape were then used for discriminant canonical variates analysis (CVA).

2.2.3. DAISY analysis

DAISY uses a Lucas n -tuple-based nearest neighbour classifier augmented by plastic self-organising map extensions to group 8- or 24-bit TIFF images based on shared patterns of pixel brightness values (detailed information on the current variant of DAISY can be found elsewhere [12]). A standard procedure is used to build all DAISY training sets, and was followed for the present study. Firstly, the images were named according to their class affiliation, and appended with a number for later identification (e.g., IDS_1.tif, FDS_37.tif, etc.) and uploaded as an image library. DAISY built the training set from this library by transforming the Cartesian format TIFF files into 32 x 32 pixel grid polar thumbnails that encompass the majority of pattern variation, while reducing processing requirements. Building the training set resulted in an ordination space in which the individual classes should be ideally clustered in discrete clouds of points. The consistency of the built training set was then tested using cross-validation analysis that provided statistical output. Where the extracted images shared a strong affinity with other images in the ordination, the system was able to identify and pass them.

3. Results

3.1. Qualitative analysis

Of the original six categories, the ‘undecided option’ was not chosen by any of the raters, indicating that the raters were able to categorise each of the 167 images into the provided five pitch contour categories (Table 1). It was found that rated pitch contour category and type of speech recipient group variables were associated, and not independent of each other ($\chi^2 = 544.038$, $df = 8$, $p < 0.001$). Cramer’s V produced a value of 0.571, which indicates a strong relationship between the two variables ($p < 0.001$). Goodman and Kruskal’s Lambda was also calculated for type of speech recipient group ($\lambda = 0.338$, $p < 0.001$) and rated pitch contour category ($\lambda = 0.314$, $p < 0.001$), the result of which showed that both variables (contour shapes and speech groups) were equally and significantly predictive of each other.

Table 1: Distribution of ratings in percent for each of the five contours across IDS, FDS and ADS.

Speech groups	Contour shape				
	Bell	Complex	Rising	Falling	Level
IDS	63.7	12	4.7	3.3	16.3
ADS	1.1	0	2.3	15.5	81.1
FDS	1.5	0	6.3	13.3	78.9

Over 60% of the IDS pitch contours were characterised as bell contours, followed by 16.3% level contours (Table 1). In both ADS and FDS this result was reversed. Here, the majority of pitch contours were characterised as level contours (ADS = 81.1%; FDS = 78.9%), with less than 2% being characterised as bell contours in both conditions. Interestingly, none of the ADS or FDS pitch contours were characterised as complex contours, whereas 12% of the IDS pitch contours fell into this category. With regards to the falling contours, more ADS pitch

contours were categorised as falling contours (15.5%) than both FDS (13.3%) and IDS (3.3%). However, the difference between ADS and FDS is minimal (2.2%). In the category for rising contours, the highest frequency was achieved by the FDS contours (6.3%) followed by IDS contours (4.7%), with ADS contours achieving the lowest frequency (2.3%).

3.2. Eigenshape analysis

Almost 70% of the shape variation in all three conditions was attributable to the first eigenshape axis. As such all outlines exhibit a fundamental similarity with each other. Separate analysis of each of the three groups demonstrated that most of the variation on the second eigenshape axis is indeed due to IDS. ADS and FDS pitch contour variation is due entirely to the first eigenshape axis, therefore indicating that these two groups are similar in that they possess very little shape variation. In contrast IDS variation is due to both the first and the second eigenshape axis and, as such, indicates wider variation across the IDS shape space. EA also provides coordinates for the graphical representation of mean shapes (Figure 1). The IDS mean eigenshape is characterised by a more exaggerated curve than either ADS or FDS. In simple terms, the shape could be characterised as a bell curve, but the top of the bell curve is flattened due to shape variation within the IDS pitch contours. The most interesting finding is that the mean shape of ADS and FDS are almost identical.

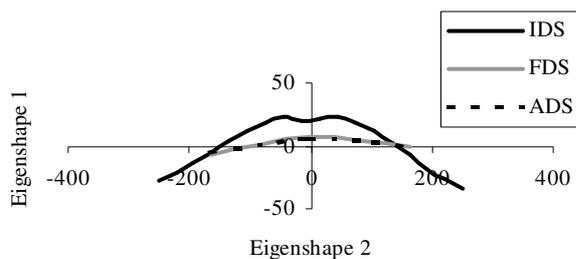


Figure 1: Comparison of mean shapes for IDS, ADS and FDS plotted on Eigenshape 1 versus Eigenshape 2.

A discriminate canonical variates analysis (CVA; simultaneous entry) was performed with speech recipient groups as the independent variable, and the scores on the 16 axes derived from EA as the dependent variable. Univariate ANOVAS revealed that the three groups differed significantly on the eigenscores of axis 1 ($F_{(2, 164)} = 66.397, p < 0.001$) and axis 10 ($F_{(2, 164)} = 66.397, p < 0.001$). Two discriminant functions were calculated. The values of the first of these functions were significantly different for IDS, ADS and FDS ($\chi^2 = 130.228, df = 32, p < 0.001, Wilks Lamda = 0.453$), whereas the second function was found not to differ significantly. The correlation between predictor variables and the discriminant functions suggested that both axis 1 (on discriminant function 1) and axis 10 (on function 2) would be the best predictors for membership of future pitch contours. Overall the discriminant functions successfully predicted outcome for 70.7% of all cases, with accurate predictions for IDS of 75.0%, for ADS of 77.0% and for FDS of 59.3%. Figure 2 shows plotting of each of the pitch contours for IDS, ADS and FDS on functions one and two. IDS is distinct from both ADS and FDS. FDS and ADS share more of their variance, but still represent discrete groupings, which is confirmed by the position of the group centroids.

In this EA it is clear that IDS is a variable group that is nevertheless distinct from ADS/FDS. These adult conditions

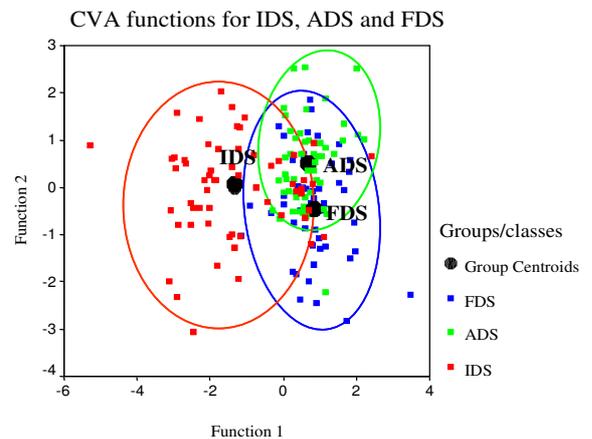


Figure 2: Combined plot of CVA functions for IDS, ADS and FDS.

exhibit very little shape variability, but do provide notable characteristics that separate one from the other. The mean ADS and FDS eigenshapes confirm this similarity, indicating that the separation relates to a shape 'tendency' rather than to a strong shape characterisation, as is the case with the mean IDS bell contour.

3.3. DAISY

Each of the classes (IDS, ADS & FDS) was recognised by DAISY as a discrete group, but the higher number of nearest neighbours in IDS (mean pass = 8) indicates a tighter clustering of the IDS data set. Note that although ADS and FDS share the same mean pass coordination rate of 8, this is an indication of clustering consistency within each class, rather than an indication of a similar clustering between these two classes. ADS and FDS images are more homogeneously spaced within their class spaces.

The individual results for each class revealed that IDS achieved the highest pass rate, as 80% of the images were correctly recognised by DAISY as belonging to the IDS class (Figure 3).

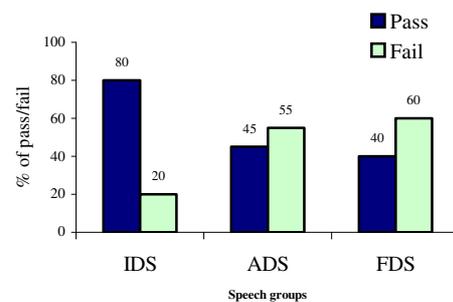


Figure 3: Comparison of pass/fail rates for IDS, ADS and FDS as provided by DAISY.

This means that those IDS images exhibited an inherent similarity to each other. Only 20% of the IDS images were rejected by the system as not belonging to the IDS class. In contrast, in ADS 45% of the ADS images were recognised as belonging to the same (ADS) class, whereas 55% were rejected. This result is to a certain degree repeated in FDS. Here, 40% of the images/pitch contours were recognised as belonging to the FDS class, whereas 60% were rejected. This association between the type of speech recipient and pass and fail rates was found to be significant ($\chi^2 = 21.616, df = 2, p < 0.001$). Pass rates of below 50% show the presence of discrete groupings, but indicate the interdigitation of class clusters; the

IDS cluster was clearly better separated from ADS and FDS than the two adult classes from each other. Fail rates indicate that most of the IDS failed images were mistaken for FDS (75%), whilst both FDS (84%) and ADS (90%) were mistaken for each other.

4. Discussion

4.1. Function of IDS pitch contours

This study confirmed the findings of previous research [e.g. 2] that pitch contours in IDS are noticeably different from those of speech directed to adults. The pitch contours of FDS were found to be more similar to ADS, which is consistent with earlier findings [8] that utilised imaginary scenarios of FDS. Contrary to research comparing tonal IDS and FDS [7], we found no evidence of linguistic exaggeration of pitch contours in non-tonal FDS. The qualitative analysis found that 75.7% of IDS consisted of bell and complex contours. The majority of ADS and FDS contours were characterised as level shapes, with almost no occurrence of bell or complex contours. Level contours are therefore characteristic of ADS and FDS, whereas bell and complex contours seem to be indicative of IDS. Although the two algorithmic approaches evaluated here do not provide named classes for the contour shapes (a fundamental requirement for the raters), the results of those techniques are in close agreement with the qualitative analysis.

An association between these exaggerated IDS shapes and particular emotional and turn taking interactions has already been noted, and it was suggested that they were responsible for the melodic quality of IDS [3]. Our findings are consistent with this viewpoint. They also naturally lead to the conclusion that, in non-tonal languages when talking to an infant, the speaker primarily uses pitch contour exaggeration to convey emotion or gain attention rather than as a tool for language transference. However, the results of the qualitative analysis also indicate a tendency for rising contours to form part of the characteristic shape space of FDS compared with ADS. This slight but tantalising finding might provide information as to how the two algorithmic approaches were able to separate ADS from FDS. Rising contours might be associated with questioning, possibly in the context of comprehension. The reliability and implications of this observation obviously require further investigation.

Because we investigated the importance of the linguistic component of IDS by comparing it to a linguistic condition (FDS), our results are not useful for discussion of the relative independence of the emotional and attentional components. An emotional speech recipient group such as partner- or pet-directed speech [e.g. 4] could potentially be more informative about the interdependence of these two components.

4.2. Evaluation of novel approaches

The results of all three approaches were consistent with each other, providing support to our overall conclusions. Each approach contributed towards the findings in different ways; the qualitative technique provided a useful insight into the proportional distribution of the contours, while the proportion of fails in each DAISY class allowed insight into the similarities between IDS and FDS. Eigenshape Analysis provides a readily useable statistical and visual representation of the groupings. Based on the present findings, we suggest that both approaches are viable options for pitch contour analysis, providing a level of objectivity that is independent of human raters. However, we suspect that such analyses would benefit from initial qualitative observations. In the present study we investigated pitch contours of words rather than

sentences, and the utility of the techniques for sentences remains to be evaluated.

5. Conclusions

We conclude that the main function of pitch contours in non-tonal IDS is emotional-attentional rather than a device to highlight important linguistic features. Furthermore, we suggest that the algorithmic approaches evaluated here present considerable potential for future pitch contour analysis.

This research was funded by the Economic and Social Research Council (ESRC) PhD studentship.

Additional material:

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

6. References

- [1] Stern, D.; Spieker, S.; Barnett, R.; MacKain, K., 1983. The prosody of maternal speech: Infant age and context related changes. *Journal of Child Language*, 10, 1-15.
- [2] Fernald, A.; Simon, T., 1984. Expanded intonation contours in mother's speech to newborns. *Developmental Psychology*, 20, 104-113.
- [3] Trainor, L.; Austin, C.; Desjardins, R., 2000. Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychological Science*, 11(3), 188-195.
- [4] Burnham, D.; Kitamura, C.; Vollmer-Conna, U., 2002. What's new pussycat? On talking to babies and animals. *Science*, 296, 1095.
- [5] Kuhl, P. K., 2000. Language, mind, and brain: experience alters perception. In *The new cognitive neuroscience*, Gazzaniga, M. S. (ed.). Cambridge: MIT Press, 99-115.
- [6] Knoll, M.; Uther, M., 2004. "Motherese" and "Chin-ese": evidence of acoustic changes in speech directed at infants and foreigners. *Jour. Acoust. Soc. Am.*, 116(4), 2522.
- [7] Papousek, M.; Hwang, S-F., 1991. Tone and intonation in Mandarin babytalk to presyllabic infants: Comparison with registers of adult conversation and foreign language instruction. *Applied Psycholinguistics*, 12, 481-504.
- [8] Biersack, S.; Kempe, V.; Knapton, L., 2005. Fine-Tuning Speech Registers: A Comparison of the Prosodic Features of Child-Directed and Foreigner-Directed. *Speech. Proceedings 9th European Conference on Speech Communication and Technology*, Lisbon.
- [9] Moore, C. A.; Cohn, J. F.; Katz, G. S., 1994. Quantitative description and differentiation of fundamental frequency contours. *Computer Speech & Language*, 8, 385-404.
- [10] Tian, J.; Nurminen, J., 2004. On Analysis of Eigenpitch in Mandarin Chinese. *Int. Conference on Spoken Language Processing, ICSLP*, Jeju Island, S. Korea.
- [11] MacLeod, N., 1999. Generalizing and extending the eigenshape method of shape space visualisation and analysis. *Paleobiology*, 25(1), 107-138.
- [12] MacLeod, N.; O'Neill, M.; Walsh, S., In press. A comparison between morphometric and artificial neural net approaches to the automated species-recognition problem in systematics. In *Biodiversity Databases: from Cottage Industry to Industrial Network*, Curry, G.; Humphries, C., (eds.) London: Taylor & Francis.
- [13] Boersma, P.; Weenink, D., 2004. *Praat: doing phonetics by computer* (Version 4.1.19) [Computer programme]. Retrieved June 2004 from <http://www.praat.org/>.
- [14] Rohlf, F. J., 2005. *tpsDig, digitize landmarks and outlines* (Version 2.0) [Computer program]. Retrieved June 2005 from <http://www.life.bio.sunysb.edu/morph/>.