

Fine-Tuning Speech Registers: A Comparison of the Prosodic Features of Child-Directed and Foreigner-Directed Speech

Sonja Biersack, Vera Kempe, Lorna Knapton

Department of Psychology
University of Stirling, Stirling, United Kingdom

sb24@stir.ac.uk, vk2@stir.ac.uk, ljk04@students.stir.ac.uk

Abstract

The present study compares prosodic features of child-directed speech (CDS) and foreigner-directed speech (FDS), to examine whether FDS is a derivative of CDS as suggested in sociolinguistic studies. Twelve female speakers completed a simple referential communication task addressed to an imaginary adult, an imaginary foreigner, and an imaginary child.

The results showed that, compared to the adult-directed baseline (ADS), participants increased pitch range and f_0 maxima when addressing a child, but not when addressing a foreigner. Furthermore, participants lowered their speech rate when addressing interlocutors with limited linguistic capacity, but did so differentially: Participants tended to lengthen pauses when addressing an imaginary foreigner, and tended to lengthen segments when addressing an imaginary child.

These findings suggest that the prosodic features of CDS and FDS are different, and that speakers have acquired knowledge about how to fine-tune their prosodic adjustments to the specific needs of different interlocutors.

1. Introduction

Research on different speech registers, so far mainly the domain of sociolinguistics [1], has focussed on the linguistic and pragmatic features of speech accommodation. In this context, researchers have emphasised the similarities between speech addressed to small children (motherese, baby talk or child-directed speech, henceforth CDS) and speech addressed to adult interlocutors with limited cognitive or linguistic capacity (elderspeak and foreigner talk or foreigner-directed speech, henceforth FDS) [2] [3], [4]. CDS, FDS, and elderspeak share such features as reduced grammatical complexity, lower propositional density, an increase of repetitions, clarifications, simplifications, and elaborations [4], [5]. However, one of the main characteristics of CDS is its unique prosody. In this study, we compare the prosodic features of CDS and FDS in order to examine whether the observed similarities between these speech registers hold in the prosodic domain.

The unique prosodic features of CDS comprise of elevated pitch, a wider pitch range and slower speech rate (e.g. [6]). They are thought to have evolved in order to manipulate infant attention and arousal [7], in particular when physical distances have to be bridged [8]. Moreover, there is evidence that the prosodic features of CDS facilitate the task of language learning by providing cues for syntactic constituents [9].

FDS is often described as being a secondary or extended form of CDS [10], not only in that it reflects the aspect of facilitating language learning, but also in that it can convey a certain notion of the listener being linguistically inferior to the speaker

[4]. However, FDS seems to be governed more by feedback from the listener than by the speaker's initial expectations [11], compared to CDS. This may reflect the huge variability in the linguistic capacity in foreign interlocutors.

To date, no studies have yet explored the prosodic characteristics of FDS. We are interested to examine to what extent CDS and FDS share commonalities in prosody, since exaggerated prosody is the most prominent feature of CDS. We investigate how speakers' general expectations about the interlocutor affect prosodic adjustment, by using a referential communication task directed to an *imaginary* addressee. This methodology allows us to control potential effects of addressee feedback, and thereby to explore baseline varieties of CDS and FDS. Our examination of prosodic features focuses on modifications of pitch range and speech rate, the two parameters that have received most attention in research on CDS.

2. Method

2.1. Participants

Twelve female native speakers of English, aged 19 to 23 years (mean age 21 years), provided three speech samples each for this study. Participants were recruited from the student population at Stirling University. Even though graduating at a Scottish university, all participants had grown up in England. None of the participants had any known history of speech impairments.

2.2. Procedure

Each participant took part in three individual sessions which were recorded with an interval of one day in between each. During each session, participants were given a simplified map containing five landmarks and a simple route, and were instructed to provide a description so that the route could be re-drawn by a potential listener, given the map with just the landmarks on it. The participants were instructed to begin each description with the following two calibration sentences "*Okay I'm now going to describe the route to you. Please draw the route exactly as I say*".

In each session, participants were given a different instruction regarding the potential listener. In the adult-directed (ADS) condition, participants were instructed to describe the route as if speaking to an adult friend or acquaintance. In the child-directed (CDS) condition, participants were instructed to describe the route as if speaking to a two-year old child. In the foreigner-directed (FDS) condition, participants were instructed to describe the route as if speaking to a fellow student who has recently arrived from a foreign country, and has just begun to learn English.

In order to avoid repetitiveness in the descriptions, three map versions with different landmarks (all simple objects denoted by mono-syllabic nouns) were created. Three further versions consisted of mirror images of the maps, resulting in a total of six different map versions which were identical in complexity but different in semantic content. The order of presentation of map versions as well as the order of imaginary addressee, was counterbalanced across speakers and sessions. The recordings were conducted using a Marantz P100 tape recorder and a uni-directional JSH MUD-805 headset microphone, and then digitised at a sampling rate of 16 kHz.

2.3. Measurements

All measurements of the prosodic features were obtained using PRAAT [12].

2.3.1. Pitch range and f_0 maxima

F_0 maxima (f_0max) were obtained by hand-labelling and measuring the peak of the f_0 contour within a phrase. To determine the pitch range the f_0 minimum was hand-labelled and measured as well, and the difference between the f_0 maximum and the f_0 minimum was converted to semitones. The measurements were performed in three instances of each map description (henceforth position): (a) in the second calibration sentence (“Please draw the route exactly as I say”), (b) in the first intonational phrase of the free description part, (usually something like: “Start at the bird...”), and (c) in the last intonational phrase of the free description part (usually something like: “...and finish underneath the rope.”). These positions were chosen because they were the most similar, or, in the case of the second calibration sentence, identical across speakers.

2.3.2. Speech rate

Speech rate was estimated in two ways: First, a coarse measure of speech rate was obtained by computing the number of syllables per second in the three positions: in the calibration sentence, in the first and in the last intonational phrase.

Secondly, since speech rate modifications can be achieved by various means such as increasing pause duration, increasing segment duration or both, a finer measure of speech rate was obtained by measuring vowel durations and pause durations in the second calibration sentence “Please draw the route exactly as I say.”, which was chosen because it was invariant, and therefore comparable, across all speakers. Specifically, we measured the vowel durations of /i:/, /ɔ:/, /u:/, /æ/ and /e/ occurring in this sentence.

3. Results

3.1. Pitch range and f_0 maxima modifications

A 3 (position: second calibration sentence, first intonational phrase, last intonational phrase) x 3 (register: ADS, FDS, CDS) ANOVA, with pitch range in semitones as the dependent variable, revealed a main effect of position $F(2,22) = 30.0$, $p < .01$. Post-hoc analysis showed that all three positions were significantly different from each other, indicating that f_0 maxima were highest in the first intonational phrase and lowest in the last intonational phrase. The effect of register fell short of significance $F(2,22) = 3.3$, $p = .06$. The post-hoc analysis revealed that the pitch range was higher in CDS than in FDS, $p < .05$ (see figure 1). This suggests that participants increased their pitch range only when speaking to an imaginary child, but not when

speaking to an imaginary foreigner. If anything, pitch ranges were slightly narrower in FDS compared to ADS, although this trend did not reach significance.

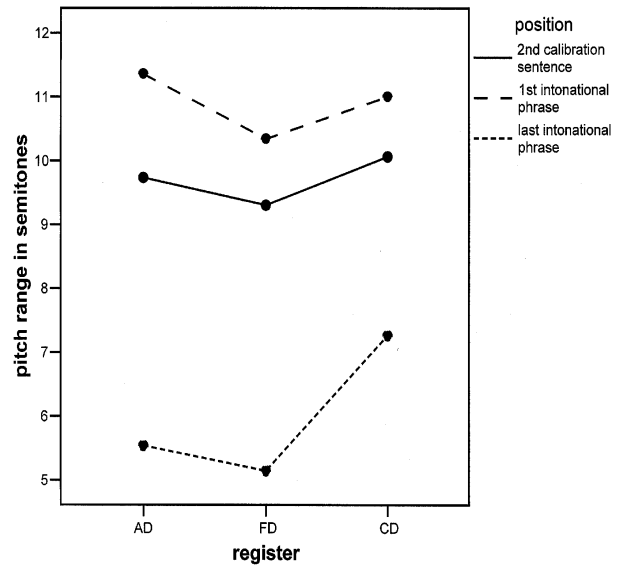


Figure 1: Pitch range (in semitones) as a function of position and speech register

A 3 (position) x 3 (register) ANOVA with f_0max as the dependent variable revealed a main effect of position $F(2,22) = 25.8$, $p < .001$. Post-hoc analysis showed that all three positions were significantly different from each other, with f_0max being highest in the first intonational phrase and lowest in the last intonational phrase. There was also a main effect of register $F(2,22) = 3.7$, $p < .05$. Post-hoc tests revealed that f_0max was significantly higher in CDS than in FDS, $p < .01$, confirming that only CDS, but not FDS, was characterised by increased pitch peaks. The difference between CDS and ADS fell short of significance, $p = 0.09$, (see figure 2) confirming the trend of lower pitch excursions in FDS compared to ADS.

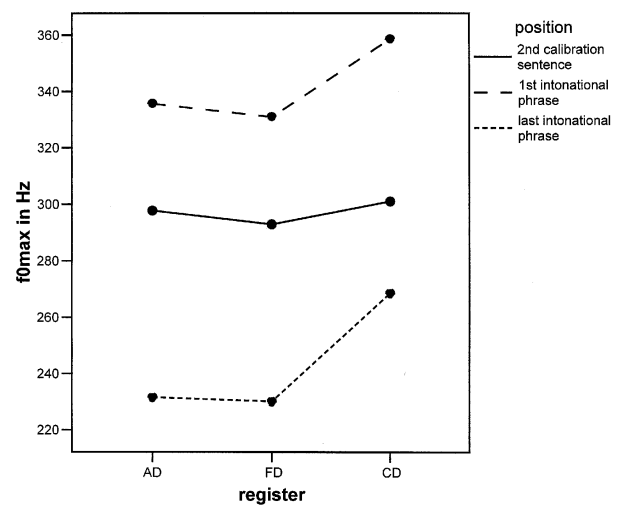


Figure 2: f_0max as a function of position and speech register

3.2. Speech rate modifications

A 3 (position: second calibration sentence, first intonational phrase, last intonational phrase) x 3 (register: ADS, FDS, CDS) ANOVA, with speech rate (in syllables per second) as the dependent variable, revealed a main effect of position, $F(2,22) = 41.2$, $p < .001$. Post-hoc tests revealed that speech rate was slowest in the second calibration sentence and fastest in the last intonational phrase. Most importantly, we observed a main effect of speech register $F(2,22) = 9.8$, $p < .001$. Post-hoc tests indicated that the speech rate was significantly slower in FDS compared to both ADS and CDS (see figure 3). There was also a tendency for the speech rate in CDS to be slower than in ADS, but this difference fell short of significance, $p = .1$. The finding that speech rate did not decrease significantly in the CDS condition is in contrast to many reported findings in the literature (e.g. [6]). This, however, is probably just an artefact of the short, in part highly automatised, phrases that were selected for comparison. When measuring speech rate in syllables per second over the entire route description, we found that overall speech rate was slower both in CDS (3.0 syll/sec) and FDS (2.8 syll/sec) when compared to ADS (3.5 syll/sec), $F(2,22) = 12.4$, $p < .001$. Post-hoc comparisons confirmed a significant difference between ADS and FDS, $p < .01$, and between ADS and CDS, $p < .01$, thereby confirming that speech rate is generally slower in CDS.

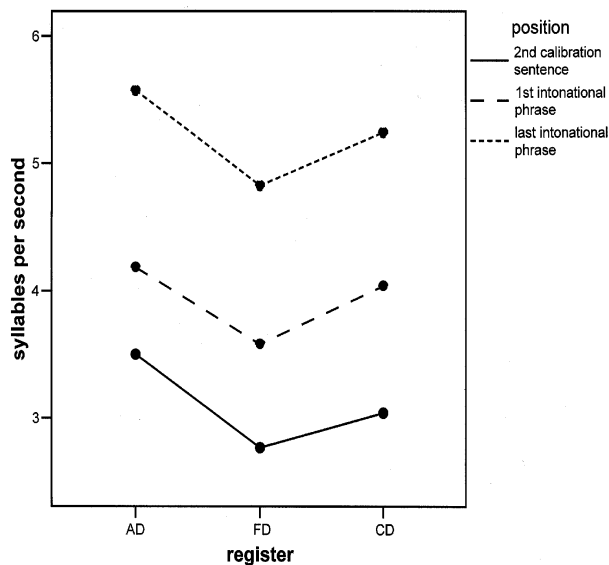


Figure 3: *Speech rate in syllables per second as a function of position and speech register*

A more detailed evaluation of speech rate modifications was conducted, to determine whether participants were lengthening pauses or segments when slowing down their speech rate. It has been shown that segment lengthening is a typical feature of CDS when expressing affection and love [13], presumably because it serves to soothe and maintain infant attention. However, when speaking to an adult who is cognitively competent but who has limited knowledge of the language, soothing and maintaining attention are not of primary importance. Instead, the speaker should attempt to give the interlocutor time to process the message. We therefore predicted that FDS should contain more pauses, while CDS should exhibit more lengthening of segments. To this end, the next set of analyses compares

mean pause durations and mean vowel durations in the second calibration sentence.

We conducted two one-way within-subject ANOVAs, with speech register as the independent variable, and mean pause duration and mean vowel duration as the dependent variables, using the second calibration sentence which was identical across speakers and speech registers. For the pause duration, the analysis yielded a main effect of register $F(2,22) = 5.8$, $p < .05$. Post-hoc comparisons revealed that the pause duration in FDS was longer than in ADS, $p < .01$; the difference between FDS and CDS fell short of significance, $p = .06$. This indicates that speakers show a strong tendency towards longer pauses in FDS speech than in the other two speech registers (see figure 4).

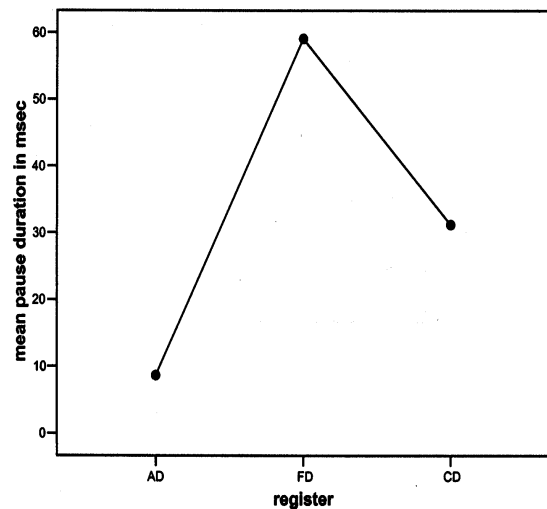


Figure 4: *Mean pause duration in the second calibration sentence across the three speech registers*

For the vowel durations, the effect of register fell short of significance $F(2,22) = 3.0$, $p = .07$. Post-hoc analysis suggested a tendency for vowels in CDS speech to be longer than in ADS speech, $p = .07$. There was also a marginal trend for somewhat longer vowels in the FDS condition, $p = .1$ (see figure 5). These analyses show that there is a strong trend towards different patterns of speech rate adjustment in FDS and CDS: As predicted, participants tended to pause longer when speaking to an imaginary foreigner, while lengthening vowels more when speaking to an imaginary child.

4. Discussion

The aim of the study was to compare prosodic features of speech addressed to an imaginary child with that addressed to an imaginary foreigner. We found that compared to an adult-directed baseline condition, participants tended to increase their pitch range, and to exhibit higher f_0 maxima in CDS but not in FDS. These changes were robust and detectable at various points of the speech episode. There was also a trend for pitch range to be slightly narrower in FDS. All of this demonstrated that speakers use attention-eliciting prosodic features appropriately when addressing an imaginary child, but omit these features when addressing an imaginary adult.

We also found that compared with ADS, participants lowered their speech rate more dramatically in FDS than in CDS. Most importantly, detailed analysis of the speech rate modifica-

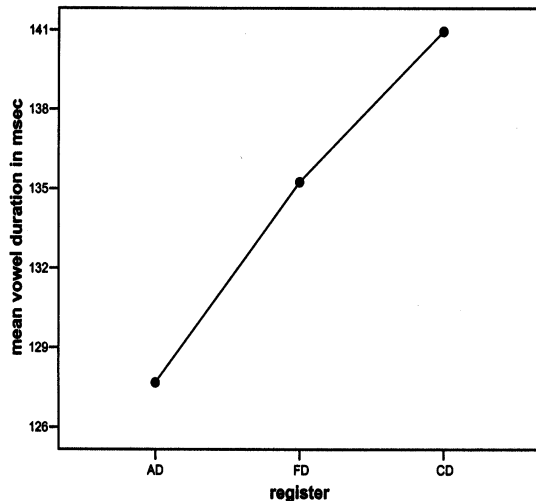


Figure 5: Mean vowel duration in the second calibration sentence across the three speech registers

tions revealed that speech rate adjustments were based on different mechanisms in CDS and FDS: Participants tended to pause more in FDS while lengthening segments in CDS, suggesting that the means by which speakers adjust their speech rate are fine-tuned towards the needs of the different interlocutors.

Taken together, these findings suggest that the prosodic features of CDS and FDS are different. Apparently, speakers are able to differentially adjust their prosody to the needs of the interlocutor, even under conditions where the interlocutor is only imaginary, and speakers have to base their adjustment on general, acquired notions about the interlocutor's needs. Presumably, these prosodic differences will be amplified in real interactions. Our results thus provide evidence against the notion that FDS is a mere derivative of CDS, and show that speakers instead fine-tune the prosodic features of their speech. The ability to do so when addressing an imaginary interlocutor suggests that some aspects of prosodic adjustment are learned, and can be invoked without interlocutor feedback.

5. References

- [1] N. Coupland and H. Giles, "Introduction: The Communicative Contexts of Accommodation," *Language and Communication*, vol. 8, pp. 175–182, 1988.
- [2] S. Kemper, P. Ferrell, T. Harden, A. Finter-Urczyk, and C. Billington, "Use of Elderspeak by Young and Older Adults to Impaired and Unimpaired Listeners," *Aging, Neuropsychology and Cognition*, vol. 5, no. 1, pp. 43–55, 1998.
- [3] N. Coupland, H. Giles, and W. Benn, "Language, Communication and the Blind," *Journal of Language and Social Psychology*, vol. 51, no. 1, pp. 53–62, 1986.
- [4] B. DePaulo and L. Coleman, "Talking to Children, Foreigners, and Retarded Adults," *Journal of Personality and Social Psychology*, vol. 51, no. 5, pp. 945–959, 1986.
- [5] S. Kemper, D. Vandeputte, K. Rice, H. Cheung, and J. Gubarchuk, "Speech Adjustments to Aging During a Referential Communication Task," *Journal of Language and Social Psychology*, vol. 14, pp. 40–59, 1995.
- [6] A. Fernald, T. Taeschner, J. Dunn, M. Papousek, B. Boysson-Bardies, and I. Fukui, "A Cross-Language Study of Prosodic Modifications in Mothers' and Fathers' Speech to Preverbal Infants," *Journal of Child Language*, vol. 16, pp. 477–501, 1989.
- [7] A. Fernald, "Human Maternal Vocalizations To Infants As Biologically Relevant Signals: An Evolutionary Perspective," in *Language Acquisition*, P. Bloom, Ed. MIT press, 1993, pp. 51–94.
- [8] D. Falk, "Prelinguistic Evolution in Early Hominins: Whence Motherese?" *Behavioral and Brain Sciences*, vol. 27, no. 4, pp. 491–541, 2004.
- [9] J. Morgan and C. Demuth, *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*. Hillsdale: Lawrence Erlbaum Associates, 1996.
- [10] C. Ferguson, "Baby Talk as a Simplified Register," in *Talking to Children*, C. Snow and C. Ferguson, Eds. Cambridge University Press, 1977, pp. 209–235.
- [11] A. Warren-Leubecker and J. N. Bohannon, "The Effects of Expectation and Feedback on Speech to Foreigners," *Journal of Psycholinguistic Research*, vol. 11, pp. 207–215, 1982.
- [12] P. Boersma and D. Weenink, "Praat: Doing Phonetics by Computer (Version 4.3.04) [Computer Program]," 2005. [Online]. Available: <http://www.praat.org>
- [13] L. Trainor, C. Austin, and R. Desjardins, "Is Infant-Directed Speech Prosody a Result of the Vocal Expression of Emotion?" *Psychological Science*, vol. 11, no. 3, May 2000.