PHONETIC REDUCTION IN SPONTANEOUS SPEECH BY CHILDREN AGED 9-14 YEARS

Outi Tuomainen, Charlotte Lee, Sonia Granlund, Valerie Hazan

Speech, Hearing and Phonetic Sciences, UCL (University College London)
o.tuomainen@ucl.ac.uk

ABSTRACT

The aim of our study was to investigate whether children distinguish between 'new' and 'given' information via phonetic reduction in spontaneous speech in a similar way to adults. An interactive 'spot the difference' game was used to elicit spontaneous speech. Word duration, fundamental frequency and vowel formant frequencies in repeated content words relative to when they were mentioned for the first time were analysed in 96 children between 9-14 years of age. There were significant developmental changes in the three acoustic-phonetic parameters between children and adults. Children produced longer words, had higher median pitch and vowel formant values than adults. However, despite these differences in spontaneous speech between children and adults, we report that, by 9 years of age (and possibly earlier), children produce phonetic reduction to highlight 'new/given' information distinction in spontaneous speech dialogues in an adult-like manner.

Keywords: spontaneous speech, speech adaptation, second mention reduction, hypo-articulation.

1. INTRODUCTION

An important aspect of speech communication is the ability to make dynamic adjustments to one's speech according to the situation in which communication is taking place. Lindblom's Hyper-Hypo model of speech production [11] states that in order to maintain success in speech communication (i.e., by producing intelligible speech and satisfying listeners' needs whilst minimising speaking effort), people tend to either 'hyper-articulate' or 'hypo-articulate' depending on the communicative situation. Hyper-articulation occurs, for example, when the talker encounters a communication barrier (e.g., background noise). It involves articulating words more clearly than when they are normally produced and it is associated with various acoustic-phonetic features of enhanced speaker effort such as longer durations and larger vowel spaces [13]. In contrast, hypo-articulation can occur, for example, when speech is contextually predictable (e.g., 'new' versus 'given' information in discourse). It involves pronouncing words less clearly and it is associated with reduction in speaking effort resulting in acoustic-phonetic adaptations such as shorter segment durations, lower pitch, reduced vowel space areas and phoneme elision [1,2,5,8]. These adaptations to word repetitions also have perceptual consequences for the listener in that when speakers repeat a word in a discourse, the repeated item tends to be less intelligible than first mentions [5].

Phonetic reduction of a repeated word (or Second Mention Reduction, SMR) in a discourse is a relatively robust phenomenon in adult talkers and it is thought to be guided both by the need to minimize speaking effort and to accentuate the distinction between new and given information [5]. In addition to the 'local' probability of a word in discourse (second mention), SMR is also mediated by other probabilistic factors such as word's lexical frequency as well as prosodic factors such as phrasal stress [1,2].

SMR can be observed across different speaking styles [3] and also in both spontaneous and read speech [3,8]. To date, however, a great majority of research on SMR has focused on adults and very little is known how this ability develops in children.

In order to fill this gap in literature, the present study examined SMR for repeated content words in children between 9-14 years of age. Previous research on speech production in children has shown that acoustic-phonetic features of spontaneous speech continue to develop well into the second decade of life [9,10]. For example, fundamental frequency characteristics of children's speech undergoes significant changes in adolescence, and their vowel space areas are significantly larger than adults' even up to 14 years of age [12]. However, despite these differences, there is evidence that children are able to produce hyper-speech and make some adult-like acoustic-phonetic adaptations to their speech in challenging listening conditions for the benefit of their interlocutor [9,14]. Moreover, it has been shown that children as young as 4 years of age are sensitive to discourse distinction between 'new/given' information [17]. Children mark new information with increased pitch and intensity in an adult-like manner but, unlike adults, they do not increase word duration to signal acoustic prominence [16,17].

The results from the above studies show that children are able to adapt their speech for the benefit of the listener by hyper-articulating in challenging conditions and placing acoustic emphasis on ‘new’ information. However, no study has specifically investigated if children minimise speaker effort and distinguish ‘new/given’ information via phonetic reduction in spontaneous speech.

Here we investigate SMR in children between 9-14 years of age and in a group of adults. We measured word duration, fundamental frequency and vowel formant frequencies in repeated content words relative to when they are mentioned for the first time in children’s spontaneous speech dialogues. We predict that children produce SMR evidenced by shorter word durations for repeated words. However, due to on-going developmental changes in pitch characteristics of children’s speech and their vowel formant frequencies [9,10,12], we predict that children differ from adults in adaptations involving pitch and vowel formants.

2. METHOD

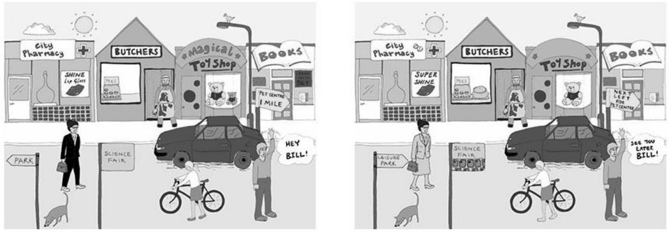
2.1. Participants

Speech was recorded from 48 single-sex pairs of Native Southern English talkers between the ages of 9 and 14 years (46 M and 50 F). The pairs knew each other and were friends. Children were divided into three age bands: thirty 9-10 year olds (14 F, 16 M), twenty-four 11-12 year olds (16 F, 8 M) and forty-two 13-14 year olds (20 F, 22 M), and their data were compared to that of 20 adults (11M, 9F, mean age: 22;04, range 18;0-29;0 years) from the LUCID corpus [4]. Adults were recorded in the same laboratory and using the same task. Participants reported no history of speech, hearing or language impairments and they passed a hearing screen at 25 dB HL or better at octave frequencies between 250 and 8000 Hz in both ears.

2.1. Procedure

During the recording, two participants sat in different rooms and communicated via headsets fitted with a condenser cardioid microphone (Beyerdynamic DT297) whilst playing an interactive 'spot the difference’ game (diapixUK; [4]) to elicit spontaneous speech. Each participant was given a different version of the same picture-scene (version A and B, see Fig. 1), and were told the pictures contained 12 differences which they had to find. They were given 10 minutes to find these differences. The speech of each participant was recorded on a separate channel at a sampling rate of 44 100 Hz (16 bit) using an EMU 0404 USB audio interface and Adobe Audition.

Figure 1: A picture pair for Diapix.



Street 3A Street 3B

2.2. Data processing

For all recordings, each channel was transcribed using freeware transcription software from Northwestern University’s Linguistics Department (Wavescroller) to a set of transcription guidelines based on those used by [15]. Word- and phoneme-level alignment software that was developed in-house was used to automatically align the transcriptions and create Praat Textgrids with separate word and phoneme (vowel) tiers which were hand-checked [6].

A list of 37 keywords consisting of the most common words featured in the diapix pictures was created. All items were high frequency [7] monosyllabic CVC/CCVC nouns, verbs and adjectives containing only monophthongal vowels (e.g., *saw*, *sea*, *dog*). Moreover, due to the nature of the task, the selected keywords had high contextual predictability.

A speaker’s first production (‘first mention’) and the following repetition of the same word later in the dialogue (‘second mention’) were extracted with an in-house Praat script (9-10 years: N=130; 11-12 years: N=98; 13-14 years: N=166, Adults: N=65 word pairs). The specific selection criteria for these items were: the items did not occur in a repair sequence (i.e., following a clarification request), did not contain laughter or any other dysfluency, and they had to occur within 2 minutes from each other. In addition to these automated selection criteria, all items were manually checked to ensure that the two mentions refer to the same entity. Because we were investigating spontaneous speech, the position of the target word in a sentence was not controlled for (a typical context for a pair: “*Next to the tree there’s a dog. The dog is sitting*.”). Word durations and median pitch (in semitones relative to 1 Hz) for all words, and vowel formant frequencies (F1,F2; Hz) for vowel /i:/ (the most frequently produced vowel; only vowel /i:/ yielded enough items for statistical analyses, 9-10 years: N=21; 11-12 years: N=21; 13-14 years: N=34, Adults: N=16 items) were extracted for first and second mention of each word. Formant frequencies were extracted from the mid-point of the vowel using Praat's formant-tracking algorithm (Burg) and checked manually.

3. RESULTS

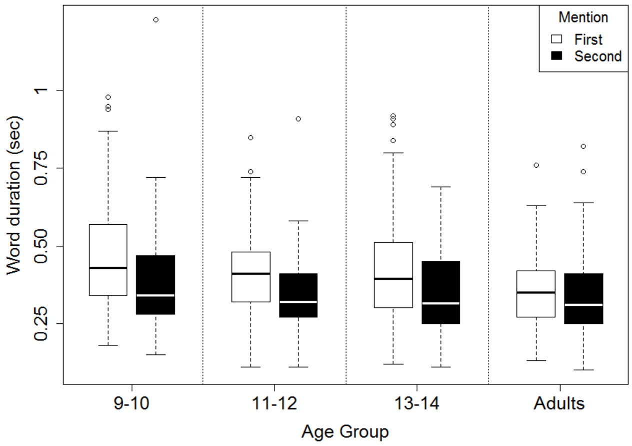
3.1. Statistics

Statistical analysis of the data was based on linear mixed-effects modelling using the lme function in the nlme package for R (R Development Core Team, 2008). The best-fitting model for each individual analysis was chosen with hierarchical approaches, that is, adding one predictor at a time to a baseline model that includes no predictors other than the intercept. The analyses were done separately for word duration, median pitch, and the F1 and F2 frequencies for vowel /i:/ of the first and second mentions, with Sex (2: Female, Male), Age band (4: 9-10 years, 11-12 years, 13-14 years, Adults) and Mention (2: first mention, second mention) as fixed effects and Participant and item as random effects. Orthogonal planned contrasts were used to evaluate if the dependent variables differed between the three child age bands and adults (i.e., three contrasts: 9-10 year olds versus adults; 11-12 year olds versus adults and 13-14 year olds versus adults).

3.2 Word duration

There were significant main effects of Age band ($\chi 2(3) = 10.60$ p=.014) and Mention ($\chi 2(1) = 71.29$ $p<.001$). No other significant main effects or interactions were found ($p>.2$). The results show that word durations (regardless of mention) decreased with increasing age (see Fig 2). However, only 9-10 year old children were significantly different from adults ($p<.001$; other comparisons $p>.05$). Moreover, we found that word duration was significantly shorter in all participants (regardless of age) when the word was mentioned the second time (First: M=0.43 seconds, SD=0.16; Second: M=0.36 seconds, SD=0.13).

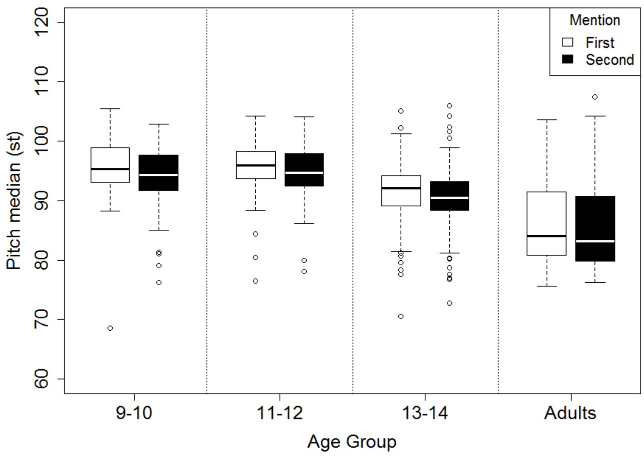
Figure 2: Word duration (seconds) for first and second mention items in children and adults.



3.3 Median pitch

There was a significant main effect of Age band ($\chi 2(3) = 106.34$ $p<.001$) and Mention ($\chi 2(1) = 9.48$ p=.002), and a significant interaction between Age band and Sex ($\chi 2(3) = 41.87$ $p<.001$). The median pitch decreased with increasing age (all comparisons, $p<.001$; see Fig 3) and it was significantly lower in the repeated items compared to the first mentions (91.6 versus 92.4 semitones). Follow-up comparisons for the Mention and Sex interaction revealed that female participants showed a greater reduction of median pitch than male participants for repeated words (Female: 94.5 versus 93.4 semitones; Male: 90.4 versus 90.0 semitones).

Figure 3: Pitch median (semitones re 1 Hz) for first and second mention items in children and adults.



3.4 F1 and F2 values for vowel /i:/

There was a significant main effect of Age band ($\chi 2(3) = 22.17$ $p<.001$) in the /i:/ F1 frequency: the F1 values decreased with increasing age and the two youngest Age bands differed significantly from adults ($p<.001$) whereas the 13-14 year olds did not ($p=.268$). No other significant main effects or interactions were found for F1 ($p>.5$).

For the F2 frequency there was a significant main effect of Age band ($\chi2(3) = 26.29$ $p<.001$), Mention ($\chi2(1) = 15.48$ $p<.001$), and a significant interaction between Mention and Sex ($\chi2(1) = 6.87$ $p=.009$). Overall, the F2 value for /i:/ reduced with increasing age (all comparisons, $p<.019$) and when the word was mentioned the second time (First: 2696 Hz; Second: 2600 Hz). Follow-up tests for the significant interaction between Mention and Sex revealed that male participants lowered the F2 value more than female participants when the word was repeated (see Table 1).

Table 1: F2 values for vowel /i:/ in first and repeated mentions (values in Hz, SD in brackets).

	First mention	Second mention	Paired t-test
Female	2750 (284)	2709 (223)	$t(49)=1.15$, $p=.254$
Male	2632 (390)	2470 (370)	$t(41)=6.19$, $p<.001$

4. DISCUSSION

The aim of the study was to investigate SMR in children aged 9-14 years. We measured word duration, fundamental frequency and F1 and F2 vowel formant frequencies of repeated words relative to when they are mentioned for the first time. We found differences in the acoustic-phonetic characteristics of spontaneous speech between children and adults for these keywords. However, all three acoustic-phonetic measures showed different maturational patterns: children had significantly longer words than adults until the age of 10 years, their median pitch was significantly higher than adults' and it had not yet reached adult levels by the age of 14 years. Moreover, children differed from adults in their formant frequency values for vowel /i:/. Despite these developmental differences in casual speech, for SMR our results suggest that from 9 years of age children show adult-like reductions for repeated words: they produce shorter words upon repetition, and they also reduce the median pitch and vowel F2 frequency in vowel /i:/. Together these data indicate that children are able to make adult-like 'new/given' information distinctions in articulation despite their articulatory-motor system not being fully matured yet. This finding is in agreement with results from clear speech adaptations in children that show that children as young as 9 years of age adapt their speech articulation when encountering a communication barrier [9].

To conclude, the acoustic-phonetic adaptations for repeated words allow speakers to reduce articulatory effort and facilitate communication efficiency for the listener. These results provide further evidence that, despite on-going cognitive, motor and socio-pragmatic changes within this age range, children are able to make subtle articulatory adjustments in their communicative discourse for the benefit of both the speaker and the interlocutor. These results also show that children are able to distinguish between subtle discourse distinctions in spontaneous speech based on the 'new/given' information status of a word.

5. REFERENCES

[1] Aylett M., Turk, A. (2004). The Smooth Signal Redundancy Hypothesis: A functional explanation for relationships between redundancy, prosodic prominence duration in spontaneous speech. *Language and Speech*, 47, 31–56.

[2] Aylett M., Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America*, 119. 3048–3058.

[3] Baker, R., Bradlow, AR. (2009). Variability in word duration as a function of probability, speech style and prosody. *Language and Speech*, 52, 391–413.

[4] Baker, R., Hazan, V. (2011). DiapixUK: a task for the elicitation of spontaneous speech dialogs. *Behavioural Research Methods*, 43, 761-770.

[5] Bard, EG., Anderson, AH., Sotillo, C., Aylett, M., Doherty-Sneddon, G., Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, 42, 1-22.

[6] Boersma, P. and Weenink, D., 2014. Praat: doing phonetics by computer, version 5.3.63.

[7] British National Corpus, version 3 (BNC XML Edition) Distributed by Oxford University Computing Services on behalf of the BNC Consortium. 2007. http://www.natcorp.ox.ac.uk.

[8] Fowler, CA., Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, 26, 489–504.

[9] Hazan, V., Tuomainen, O., Pettinato, M. Submitted. Suprasegmental characteristics of spontaneous speech produced in good and

challenging communicative conditions by talkers aged 9 to 14 years old. *Journal of Speech, Language and Hearing Research*.
[10] Lee, S., Potamainos, A., Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America*, 105, 1455–1468.
[11] Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. In: Hardcastle, W. J. and Marchal, A. (eds.) *Speech Production and Speech Modelling*. Dordrecht: Kluwer Academic Publishers, 403-439.
[12] Pettinato, M., Tuomainen, O., Granlund, S., Hazan, V. Under review. Vowel space area in later childhood and adolescence: effects of age, sex and ease of communication. *Journal of Phonetics*.
[13] Smiljanic, R., Bradlow, AR. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Linguistics and Language Compass*, 3, 236–264.
[14] Syrett, K., Kawahara, S. (2013). Production and Perception of Listener-Oriented Clear Speech in Child Language. *Journal of Child Language*, 14, 1-17.
[15] van Engen, KJ., Baese-Berk, M., Baker, R., Choi, A., Kim, M., Bradlow, AR. (2010). The Wildcat Corpus of Native and Foreign-Accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech*, 53, 510-540.
[16] Watson, D., Arnold, JE., Tanenhaus, MK. (2010). Tic Tac TOE: Effects of predictability and importance on acoustic prominence in language production. *Cognition*, 114, 462-3.
[17] Wonnacott, E., Watson, D. (2008). Acoustic emphasis in four year olds. *Cognition*, 107, 1092-1101.

6. ACKNOWLEDGEMENTS

This work was supported by the Economic and Social Research Council [grant number RES-062-23-3106]. We thank all parents and children who took part in the study.