Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal

Montreal, Canada

2 - 7 June 2013

Speech Communication

Session 2pSCb: Speech Intelligibility (Poster Session)

2pSCb13. The development of clear speech strategies in 9-14 year olds

Michèle Pettinato and Valerie Hazan*

***Corresponding author's address: Speech Hearing and Phonetic Sciences, University College London, Chandler House, London, WC1N 1PF, Greater London, United Kingdom, v.hazan@ucl.ac.uk**

This study investigated the development of global clear speech strategies of child talkers. Two groups of 20 talkers aged 9-10 (children) and 13-14 (teens) were recorded in pairs while they carried out spot the difference picture tasks (diapix), either hearing each other normally (NB condition) or with one talker hearing the other via a three-channel noise vocoder (VOC condition). Acoustic-phonetic analyses focused on the talker having to overcome the communication barrier. Data were compared to those for twenty of the adults in Hazan and Baker (2011) [J. Acoust. Soc. Am., 130, 2139-2152]. The three age groups did not differ in task transaction time for NB, but children took significantly longer to complete the task in VOC than teens or adults who took equally long. Children spoke at a slower speech rate overall than teens, while teens and adults did not differ; all groups significantly reduced their speech rate in VOC relative to NB. Adults hyperarticulated vowels in VOC but children and teens showed only minor adaptations. These results suggest that although 9-10 year olds use some strategies to clarify their speech in difficult conditions, other strategies continue to develop into late adolescence.

Published by the Acoustical Society of America through the American Institute of Physics

Received 22 Jan 2013; published 2 Jun 2013

INTRODUCTION

Clear speech is a speaking style which talkers typically adopt when transmission of the speech signal is compromised, e.g. by noisy conditions, distortion or when talking to a hearing-impaired person. Clear speech can involve overall changes to speech rate, intensity and F0 as well as more fine-grained modifications at the phonetic level such as increased VOT differences or vowel hyperarticulation (see Smiljanic & Bradlow, 2009 for a review).

Speakers possess remarkable flexibility in the manner in which individual acoustic and phonetic modifications are instantiated: Hazan & Baker (2011) showed that speakers differentially enhanced aspects of F0, intensity and vowel hyperarticulation when overcoming either multi-talker babble or a 3-channel noise vocoder. Indeed, the phonetic and phonological structure of a language can influence which features are enhanced (Smiljanic & Bradlow, 2009 for a review), and bilingual speakers may use different enhancements in their respective languages (Smiljanic & Bradlow, 2008; Granlund et al., 2012).

Since adults are able to adapt their speech in this skilled manner and clear speech is in part a linguistic phenomenon, the question of how this ability develops becomes relevant. Whilst there is a body of research with adults, clear speech research with children is sparse. Redford & Neuman-Gildersleeve (2009) examined clear speech strategies in children aged 3-5 and reported that around 4 years perceivable differences between clear and casual speech styles started to emerge, although acoustic measurements showed that the manner in which this was achieved was not adult-like. Furthermore, children's speech has been found to be characterized by greater inherent variability in the articulation and phonetic realization of phonemes, which only gradually reduces to adult-like levels around early adolescence (Lee et al 1999, Gerosa et al., 2006). Children might hence experience difficulties in producing certain aspects of clear speech, as variability may affect the ability to successfully exaggerate phonemic targets.

The present study therefore examined two common measures of clear speech: speech rate and vowel hyperarticulation in two groups of young speakers (ages 9-10 and 13-14), comparing these to adults. It has been suggested that hyperarticulation is linked to the decrease in speech rate in clear speech (Moon & Lindblom, 1994; Krause & Braida, 2004). Given children’s lack of fine control over their speech production, it is particularly interesting to examine the development of these aspects of clear speech.

It was predicted that speech rate should be equally available to both groups, because it is less dependent on fine phonetic abilities. Vowel hyperarticulation was predicted to be less evident in the younger group, whereas the teenage group was expected to display more adult-like abilities, since phonetic variability has been reported to approximate adult level by this age (Lee et al., 1999; Gerosa et al., 2006).

The study used diapix (Van Engen et al., 2010; Baker & Hazan, 2011), a communicative task which can be carried out in good and poor listening conditions to elicit different speaking styles. Recent research on clear speech has moved from a reliance on read laboratory speech to investigations of more naturalistic speech, as the indices of clear speech differ between these settings (e.g., Hazan & Baker, 2011) and an interactional approach reflects the listener-oriented nature of clear speech (Lindblom, 1990).

METHOD

Participants

Three groups of 20 participants (8 males,12 females) were compared: children aged 9-10, adolescents aged 13-14 and adults aged 19-29. The adult recordings of diapix tasks were a sub-set of the LUCID corpus (Hazan & Baker, 2011).

All participants passed a hearing screen at 25dB HL or better for the range 250 – 8000Hz. A short parent questionnaire confirmed that none of the children and teens had a history of speech or language difficulties, which was also the case for the adult participants.

Procedure

In order to elicit spontaneous speech which varied in clarity, two participants (talker A and B) were seated in separate sound-treated rooms unable to see each other. They each had a similar picture and communicated via a headset and microphone to find the 12 differences between the pictures (for more details, see Baker & Hazan, 2010). The conversations were carried out in two conditions: no barrier (NB), during which talkers heard each other normally and VOC, when talker A’s voice was passed through a 3-channel noise-excited vocoder (Hazan & Baker, 2011), which meant that comprehension was reduced for talker B and A had to increase the clarity of her speech. The two talkers in a pair switched roles so that all produced speech as talker A. For each condition, there were therefore two repetitions of the task with a different picture and talker A. Recordings lasted for about 10 minutes, yielding around 4 minutes of analyzable speech for talker A once silences, fillers, non-speech sounds such as laughter and sections with background noise had been excluded.

Data Processing

The speech of talkers A and B was separately transcribed orthographically and automatically aligned at word and phoneme-level to create Praat textgrids (Boersma and Weenink, 2012). Analyses were carried out on the speech of talker A, who was expected to produce clear speech in the VOC condition.

Analyses

To ascertain whether the task difficulty was comparable for the three age groups and in the different conditions, the time taken to find the first eight 8 differences in the pictures was compared.

Mean word duration was used as a measure reflecting the average speaking rate of a talker in a given condition. The duration of each of the orthographically annotated regions of the speech recording for talker A was calculated using Xplic8 software (Tang et al., 2012). Portions of the recording containing laughter, silence, breath, hesitations, fillers and agreements were excluded from this calculation. Mean word duration was then obtained by dividing the total duration of ‘speech’-labeled regions by the number of individual words within these regions.

The phoneme-level tier in the Praat textgrids was used to identify regions of the signal containing the vowels [iː], [æ] and [ɔː]; this was done for content words only. Formant tracking algorithms in Praat were used to get formant estimates for F1 and F2. Individual vowels for which both formant estimates were outside of 3 standard deviations of the F1/F2 mean per vowel were excluded as outliers. On average, 60 [iː], 33 [æ] and 20 [ɔː] vowel tokens were included in the calculations of vowel measures per talker per condition. Formant estimates were then normalized to ERB values to reduce the effect of anatomical differences due to gender and age, and median F1/F2 frequency values were then calculated per vowel per talker. For each speaker, a measure of F1 range (in ERB) was derived by subtracting F1[iː] from F1[æ] , giving an indication of how much vowels were differentiated in terms of height. The degree to which the front/back distinction was instantiated was explored by examining the F2 range derived by subtracting F2[ɔː] from F2[iː].

RESULTS

All statistical analyses were carried out using repeated-measures ANOVAs with condition (NB, VOC) as a within-subject factor and group (adult, teen, child) as a between-subject factor. In order to establish whether the task difficulty was similar for all three groups, transaction time (i.e. the time to find the first eight differences in the pictures) was analyzed. There were significant main effects for condition [$F(1,57) = 46.59$, $p < .001$] and group [$F(2,57) = 7.24$, $p = .002$], and the interaction between these was significant [$F(2,57) = 5.56$, $p = .006$]. Post-hoc Bonferroni pairwise comparisons showed that transaction time did not differ for adults and teenagers, but that longer transaction times were obtained for the child group than for the two other groups (for means and standard deviations, see Table 1). To understand where this difference originated, two separate one-way ANOVAs with group as between-subject factor were carried out on transaction times in each condition. In the NB condition, there was no effect of group [$F(2,57) = 2.34$, $p > .05$], confirming that all groups were equally able to carry out the task. In the VOC condition, the groups differed significantly [$F(2,57) = 8.04$, $p = .001$]. Post-hoc Bonferroni pairwise comparisons revealed that whilst adults did not differ from adolescents, children were significantly slower than both adults and adolescents, suggesting that the VOC condition presented a bigger challenge to the younger group.

TABLE 1. Means and standard deviations in parentheses for each of the measures in all groups and both conditions

	No Barrier	Vocoded
	Time to reach 8 differences (seconds)	
9-10	288 (125)	552 (267)
13-14	222 (65)	351 (115)
adults	279 (110)	360 (104)
	Mean word duration (seconds)	
9-10	.310 (.03)	.420 (.10)
13-14	.282 (.02)	.351 (.05)
adults	.262 (.03)	.332 (.06)
	F1 range (ERB)	
9-10	5.95 (.58)	6.27 (.85)
13-14	5.72 (.85)	6.07 (.91)
adults	2.99 (1.18)	4.20 (.99)
	F2 range (ERB)	
9-10	8.22 (.60)	8.38 (.95)
13-14	8.03 (.84)	8.30 (1.21)
adults	2.75 (.68)	3.71 (.63)

Next, the speech produced by each talker in the NB and VOC conditions was analyzed. To test whether participants significantly slowed down their speech rate to achieve greater clarity and whether this changed between age groups, analyses were carried out on mean word durations. There were significant main effects for condition [$F(1,57) = 84.30$, $p < .001$] and group [$F(2,57) = 13.73$, $p < .001$], but no significant interaction [$F(2,57) = 2.29$, $p > .05$], with all three groups significantly increasing their mean word duration in the VOC condition. Bonferroni pairwise comparisons indicated that the children had significantly longer mean word durations than both teens and adults, but that the latter two did not differ. Reducing speech rate was therefore a strategy which was equally available to all three groups, and children spoke at a slower speech rate overall than teens and adults, possibly due to inherently slower speech rates in younger children (Omir & Grinfeld, 2011).

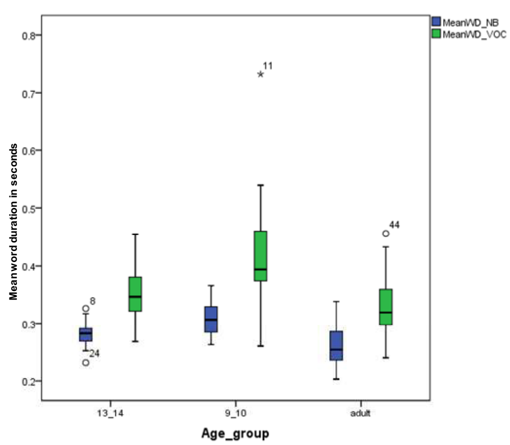

FIGURE 1. Boxplots of mean word durations (seconds) in the NB and VOC conditions for each group

Vowel hyperarticulation was investigated with separate analyses on F1 and F2 ranges (see Table 1 and Figure 2). For F1 ranges, there were main effects of condition [$F(1,57) = 38.75$, $p < .001$] and group [$F(2,57) = 57.27$, $p < .001$], as well as a significant interaction between these factors [$F(2,57) = 8.4$, $p = .001$]. Paired-samples t-tests showed that F1 ranges for adults and teens showed a significant increase in the VOC condition while children's did not. The same analysis on the F2 range showed similar results: condition was a significant effect [$F(1,57) = 33.81$, $p < .001$] as well as group [$F(2,57) = 268.2$, $p < .001$]; again the two interacted [$F(2,57) = 9.88$, $p < .001$]. Paired-samples t-tests showed that F2 range increased significantly for adults in the VOC condition but not for children and teens. As can be seen in Figure 2, adults were the only group who clearly expanded their F1 and F2 ranges in the VOC condition, whereas the child and teen groups only made minor adaptations, which suggests that these groups are not reliably hyperarticulating vowels. A better sense of the effect this is having on the vowel space can be garnered by plotting the values for each of the three vowels (see Figure 3); it can be seen that the vowel space is indeed expanded in the VOC condition for adults but much less visibly so for the child and teen groups.

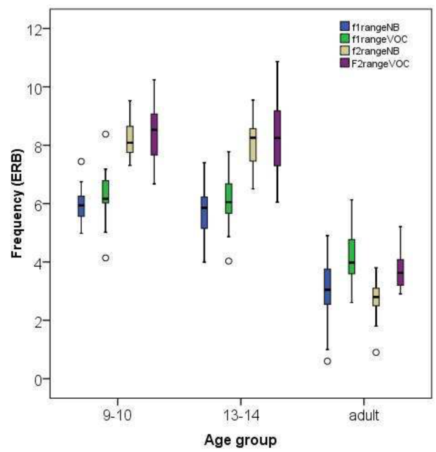

FIGURE 2. Boxplots of ERB values for F1 and F2 ranges in the NB and VOC conditions for each group. The F1 range represents the difference in F1 between [æ] and [iː]. The F2 range is the difference in F2 between [iː] and [ɔː].

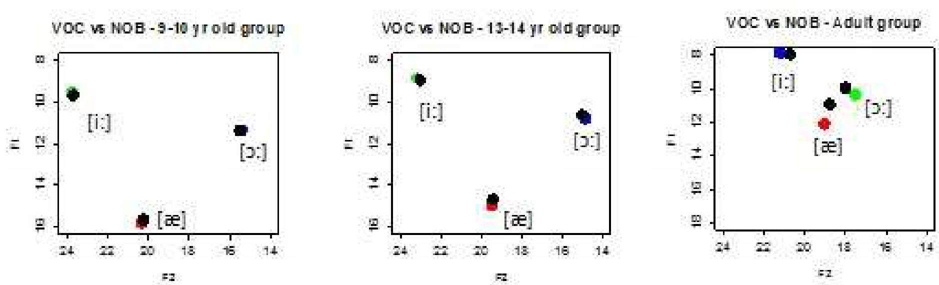

FIGURE 3. ERB values for the vowels [iː], [æ] and [ɔː] in children, teens and adults. The colored dots represent vowels in VOC and black dots are vowels in NB.

DISCUSSION

This study considered the manner in which adults, 9-10 year old children and 13-14 year old teenagers use speech rate and vowel hyperarticulation when endeavoring to overcome the effects of a 3-channel vocoder in a communicative task. To our knowledge this is the first study investigating clear speech in these younger age groups with a spontaneous speech task. It was predicted that all age groups would slow down their speech to overcome adverse listening conditions, but that hyperarticulation should be less evident in children as this relied on more mature phonetic abilities, although this strategy was expected to be available to the teen group.

It is important to note that the diapix task was essentially accessible to both children and teenagers: when there was no communication barrier, neither differed from adults in the time taken to find eight differences between two pictures. In the condition where vocoding distorted the speech of one talker, teens completed the task within the same time frame as adults, but children became slower at identifying eight differences, as the degradation of the speech signal presented an additional, possibly cognitive, challenge. They may have strained under the double impact of talker A not clarifying her speech enough to overcome the effects of vocoding, and her partner inherently needing more information in the signal: the amount of channels needed to understand vocoded speech decreases through childhood (Eisenberg et al., 2000; Nittrouer et al., 2000), and children up to teenage years are in general more affected by speech degradation (Johnson, 2000).

In spite of finding the task equally difficult, teens did not use the two enhancement strategies that we evaluated to the same extent as adults. In fact, their performance was more similar to that of the younger group: both teens and children slowed down speech rate in the VOC condition, but unlike adults, neither group provided clear evidence of vowel hyperarticulation. This partially contradicted the initial predictions, as finer phonetic enhancements remained absent in the older group. It therefore seems that the strategy of slowing down speech as a means of clarifying it precedes the strategy of hyperarticulating vowels for the same purpose. In adults, a strong link between the two has been suggested (Moon & Lindblom, 1994; Krause & Braida, 2004). The relationship between these abilities through development therefore merits further investigation. These initial results suggest that the ability to use multiple strategies to make speech adaptations for the benefit of an interlocutor in communicative speech develops late and may still be ongoing in teenage years.

ACKNOWLEDGMENTS

This work was supported by the Economic and Social Research Council [grant number: RES-062-23-3106]

REFERENCES

Amir, O., and Grinfeld, D (**2011**). “Articulation Rate in Childhood and Adolescence: Hebrew Speakers, ” Language and Speech **54**, 225-240.

Baker, R., and Hazan, V. (**2011**). “DiapixUK: a task for the elicitation of spontaneous speech dialogs,” Behav. Res. Meth. **43**, 761-770.

Boersma, P. & Weenink, D. (**2012**). “Praat: doing phonetics by computer” [Computer program]. Version 5.3.07, http://www.praat.org/. (last retrieved 5 March 2012.)

Eisenberg, L. S., Shannon, R. S., Schaefer Martinez, A., Wygonski, A., and Boothroyd, A. (**2000**) “Speech recognition with reduced spectral cues as a function of age, ” J. Acoust. Soc. Am. **107**, 2704-2710.

Gerosa, M., Giuliani, D. and Brugnara, F. (**2007**). “Acoustic variability and automatic recognition of children’s speech,” Speech Commun. **49**, 847-860.

Granlund, S., Hazan, V., and Baker, R. (**2012**). “An acoustic-phonetic comparison of the clear speaking styles of late Finnish-English bilinguals,” J. Phon. **40**, 1-12.

Hazan, V. and Baker, R. (**2011**). “Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions,” J. Acoust. Soc. Am. **130**, 2139-2152.

Johnson, C. E. (**2000**) “Children's phoneme identification in reverberation and noise,” J. Speech Lang. Hear. Res. **43**, 144-57.

Krause, J. C., and Braida, L. D. (**2002**). “Investigating alternative forms of clear speech: the effects of speaking rate and speaking mode on intelligibility,” J. Acoust. Soc. Am. **112**, 2165–2173.

Lee, S., Potamianos, A., & Narayanan, S. (**1999**). “Acoustics of children’s speech: Developmental changes of temporal and spectral parameters,” J. Acoust. Soc. Am. **105**, 1455-1468.

Lindblom, B. (**1990**) “Explaining phonetic variation: a sketch of the H&H theory”. in *Speech Production and Speech Modelling*, edited by W. J. Hardcastle and A. Marchal (The Netherlands: Kluwer Academic), pp. 403-439.

Moon, S.-J., and Lindblom, B. (**1994**). “Interaction between duration, context, and speaking style in English stressed vowels,”J. Acoust. Soc. Am., **96**, 40- 55.

Nittrouer, S., Lowenstein J.H., and Packer, R., R. (**2009**). “Children discover the spectral skeletons in their native language before the amplitude envelopes, ” J. Exp. Psychol. Hum. Percept. Perform. **35,** 1245-53.

Redford, M.A., Gildersleeve-Neumann, C.E. (**2009**). “The development of distinct speaking styles in preschool children, ” J. Speech Lang. Hear. Res. **52**, 1434-1448.

Smiljanic, R., and Bradlow, A.R. (**2008**). “Stability of temporal contrasts across speaking styles in English and Croatian, ” J. Phon., **36**, 91–113.

Smiljanic, R., and Bradlow, A. R. (**2009**). “Speaking and hearing clearly: Talker and listener factors in speaking style changes, ” Linguistics and Language Compass, **3**, 236–264.

Tang, Y., Nicolao, M., Raitio, T., Jokinen, E., Koutsogiannaki, M., Godoy, E., Granlund, S., Aubanel, V., Sfakianaki, A., Mowlaee, P., Hazan, V. & Stylianou, Y. (**2012**). “exPlic8” [Computer program]. Version 1.0

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M. and Bradlow, A. R. (**2010**). “The Wildcat Corpus of Native- and Foreign-Accented English: Communicative efficiency across conversational dyads with varying language alignment profiles, ” Lang. Speech **53**, 510-540.