



Children's timing and repair strategies for communication in adverse listening conditions

Valerie Hazan, Michèle Pettinato

Department of Speech Hearing and Phonetic Sciences, University College London, London, UK

v.hazan@ucl.ac.uk, m.pettinato@ucl.ac.uk

Abstract

This study investigated the development of clear speech strategies in children. Groups of 20 talkers aged 9-10 and 13-14 years were recorded in pairs while they carried out 'spot the difference' picture tasks, either while hearing each other normally ('no barrier') or when one talker heard the other via a noise vocoder (VOC), which led their interlocutor ('talker A') to clarify their speech to maintain effective communication. Data were compared to those previously collected for 20 adults. Mean word duration, number and duration of pauses were calculated for talker A. Strategies used in response to direct requests for clarification were also classified. Children spoke at a slower rate than teens and adults, who did not differ. Relative to 'no barrier', all groups significantly reduced their speech rate in VOC and used longer pauses, but the relative change in pause rate across conditions was greater for adults than children or teens. In response to a direct clarification request, children and teens used a higher rate of repetitions than adults who used more varied strategies. These results suggest that although children use some strategies to clarify their speech in difficult conditions, other strategies continue to develop until late adolescence.

Index Terms: speech production, speech and language acquisition.

1. Introduction

Much communication between interlocutors occurs while one or both individuals are listening in less than ideal conditions, e.g., in a background of noise, via a poor quality phone line, or when one of the talkers has a hearing impairment. Adult talkers are usually able to compensate for these potential barriers to effective communication by adopting a 'clear speaking style'. Clear speech can involve a decrease in speech rate [1,2,3], an increase in the number and duration of pauses [1,4], increases in intensity [5] and in F0 median and range [1], as well as more fine-grained modifications at the phonetic level such as vowel hyperarticulation and increased VOT differences in stop voicing contrasts (see [6] for a review). There has been much recent interest in characterizing clear speech, as understanding how human talkers can clarify aspects of their speech may inform work on automatic speech enhancement.

Much of this work has investigated the speech adaptations that talkers make when directly affected by noise [e.g. 7]. However, in our work [8], we have focused on adaptations that talkers make entirely for the benefit of their interlocutor. We do this by evaluating the changes that talkers (who are hearing normally) make when they are communicating with someone who is impaired due to a simulated cochlear implant (hearing vocoded speech), due to the presence of background noise or because they are not native speakers of English. The Hyper-Hypo (H&H) theory of speech production [9] is a useful framework for our study as it discusses how the control that

talkers have over their speech production is used to maximize communication efficiency in different communicative situations; there is an ongoing tension between a talker's desire to minimize effort and the maintenance of effective communication. We showed that talkers adjust the speech adaptations that they make in each of the three 'barrier' conditions that we examined [8]. For example, pitch characteristics were enhanced in the babble condition but not in the noise-excited vocoder (VOC) condition, where they would bring little benefit. Also, linguistic changes such as a reduction in type-to-token ratio were more prevalent in the L2 condition, where the communication barrier was linguistic, than in the babble condition when the barrier was acoustic.

Since adults are able to adapt their speech in this skilled manner and since clear speech is in part a linguistic phenomenon, the question of how strategies to deal with communication in adverse conditions develop in children becomes relevant. Clear speech research with children is sparse. Perceivable differences between clear and casual speech styles have been shown to start to emerge at around 4 years [10], and older children, while reading simple sentences, showed similar clear speech as adults in [11]. Children's speech is known to be more variable in the articulation and phonetic realization of speech sounds, and this variability only gradually reduces to adult-like levels in early adolescence [12, 13]. This lack of control over aspects of speech production may affect the ability to make specific adaptations in difficult listening conditions.

In our current study, we are investigating speech communication in good and adverse conditions between pairs of children aged 9-10 and 13-14 years, using the same diapix 'spot the difference' picture task [14] and adverse listening conditions (vocoder, babble noise) as in our previous study with adults [8]; note that we are only reporting the 'no barrier' and vocoder (VOC) conditions in this paper. In initial analyses [15], we found that children aged 9-10 years, 13-14 years and adults did not differ in the time it took them to find the first eight differences when the two interlocutors could hear each other normally ('no barrier' condition), showing that the task that we used did not tax children to a greater extent than adults. When doing the task while one of the speakers heard the other via a simulated cochlear implant (VOC condition), teens completed the task within the same time frame as adults, but children took significantly longer, suggesting that the degradation of the speech signal presented an additional, possibly cognitive, challenge for them. Children had a slower speech rate overall than teens, while teens and adults did not differ; all groups significantly reduced their speech rate in the VOC condition. Adults hyperarticulated vowels in the VOC condition, 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, such as the use of segmental enhancements, continue to develop into late adolescence.

In this paper, we further investigate the development of clear speech strategies in this group of children and teens. First, we explore another aspect of temporal organization in clear speech. The introduction of pauses, and increase in pause durations has been found in studies of clear speech to be an effective strategy for improving speech clarity [e.g., 16]. We investigate how children and teens use pausing as a strategy while communicating in good and adverse conditions.

Another important aspect of communication is how speakers deal with communication breakdowns that occur in interactions in difficult conditions. We investigate the conversational repair strategies that children and teenagers adopt in response to direct requests for clarification. Adults use a range of strategies to deal with such requests [e.g., 17]. In addition to direct repetitions of the whole of the misunderstood utterance, strategies include: expansion, rephrasing and reduction. Studies that involved sequences of requests for clarification controlled by adult experimenters while interacting with children showed that children aged 9 years used a broader range of strategies than younger children, suggesting a developmental pattern [18].

Our predictions were that some of the temporal and discourse strategies that adults use to clarify their speech will be late acquired and may be seen in the 13-14 year old teens but not in 9-10 year olds, even though most aspects of speech development are well established by that age. We predict that while reducing speech rate may be a strategy used by all age groups, the use of pausing as a strategy to improve clarity may be seen in teens and adults but not children. Similarly, we expect the 9-10 year olds to be less diverse in the strategies used for dealing with direct requests for clarification than teens and adults.

The study used diapix [19, 14], 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 clear speaking styles elicited via instructions to read written materials clearly, to investigations of more naturalistic speech. Indeed, the indices of clear speech differ between these settings [e.g., 8] and an interactional approach reflects the listener-oriented nature of clear speech [9].

2. Method

2.1. Participants

Three groups of 20 participants (8 males, 12 females) were compared: children aged 9-10 years, teens aged 13-14 years and adults aged 19-29 years. The child and teen recordings are being collected as part of an ongoing study while the adult recordings were a randomly-chosen sub-set of the LUCID corpus that used the same test materials and was recorded in the same recording conditions as the new corpus [8]. 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.

2.2. Procedure

In order to elicit spontaneous speech which varied in clarity, two participants (talker A and B), who were known to each other, were seated in separate sound-treated rooms unable to see each other. They carried out a 'spot the difference' task using the diapixUK picture materials [14]. Each talker had a coloured cartoon picture depicting a complex scene and had to communicate with their interlocutor via a headset fitted with a microphone to find the 12 differences between the two pictures (for more details, see [14]). The conversations were carried out in two conditions: 'no barrier', during which both talkers heard each other normally and VOC, when talker A's voice was passed through a 3-channel noise-excited vocoder [8]. This resulted in comprehension being reduced for talker B so that talker A had to increase the clarity of her speech in order to complete the task successfully. The two talkers in a pair switched roles so that all produced speech as talker A. In each condition, recordings lasted for about 10 minutes, yielding around 4 minutes of analyzable speech for talker A per condition once silences, fillers, non-speech sounds such as laughter and sections with background noise had been excluded. The recordings that had been made with adults in a previous study [8] followed this same experimental design. In the adult study, three picture tasks were carried out per condition, rather than one as used in the child study, but only the recording from the first picture task for the 'no barrier' and VOC conditions were used in the analysis.

2.3. Data processing

The speech produced by talkers A and B was transcribed orthographically. The transcription files were aligned to the waveform using automatic alignment software developed at UCL. The aligner created Praat TextGrid files [20] with word- and phone-level tiers for each file.

2.4. Data analysis

Temporal measures: two temporal measures were calculated. The first measure related to the number and duration of within-speaker pauses. During transcription, any within-speaker pauses which were over 500ms in duration and which were not interrupted by the interlocutor were tagged. The number of pauses and their duration for each speaker per condition was then calculated using a Praat script.

The mean word duration data, reported in [15], is also briefly reviewed. 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 Explic8 software [21]. 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.

Classification of repair strategies: We classified the repair strategies used in response to direct requests for clarification using a classification based on [17]. This classification was chosen for its simplicity and not because it is a gold standard within the literature on communication breakdowns in children. First, orthographic transcripts of the dialogues produced by all 60 speakers (as talker A) in the VOC condition were produced from the Praat textgrids. An

experimenter identified all instances of direct request for clarification (e.g., ‘what?’, ‘say that again’, ‘sorry?’). These included clarifications requested by both talkers, but most clarifications were made by talker A, who was instructed to take the lead in this condition. All instances were then classified using one of four categories: (1) repetition, when the entire utterance that preceded the request for clarification was repeated; (2) expansion, where additional information was given (e.g. Talker A: *he is saying the word ‘shoot’*, Talker B: *what?*, Talker A: *he is saying the word ‘shoot’ to the girl*); (3) rephrasing, when a semantic equivalent was given, either a single word or whole phrase (e.g., Talker A: *there is no more hay*, Talker B: *what?*, Talker A: *there is only two bits of hay*); (4) reduction, where syntactic complexity was reduced by removing part of the sentence (e.g. Talker A: *okay is there anything else around there because there’s nothing for me*, Talker B: *sorry?*, Talker A: *anything else that you’ve got?*). The number of instances where multiple turns were needed to deal with the lack of comprehension (‘stacked requests’) was also noted. Counts of each of these strategies were transferred to a spreadsheet, and proportions of each strategy as a function of the total number of instances of repetitions, reductions, expansions and rephrasings for each talker was calculated.

3. Results

3.1. Pause analysis

First, pause rate was calculated (for each talker A and for each condition) as the number of within-talker pauses divided by the number of words (including fillers but excluding interrupted words). The relative change in pause rate between ‘no barrier’ and VOC was then calculated (see Figure 1). The data violated Levene’s test of the equality of variances so non-parametric (Kruskal-Wallis) tests were run. There was a statistically significant difference between the three age groups ($\chi^2(2, N=59) = 10.60, p = 0.005$). Post-hoc tests suggest that this is due to adults showing greater relative change in pause rate than children and teens who did not differ.

Next, median pause duration was calculated (see Table 1). A repeated-measures ANOVA on the log-transformed median durations revealed a main effect of condition [$F(1,57)=6.87; p<0.02$], with longer pauses obtained in VOC (mean: 871 ms) than ‘no barrier’ (816 ms), and a main effect of age group [$F(2,57)=10.90; p<0.0001$]; the condition by age interaction was not significant. Bonferroni-adjusted pairwise comparisons showed that pauses produced by adults were shorter than those produced by children or teens, who did not differ, but this may simply reflect a difference in speaking rates.

Table 1. Means and standard deviations for median pause duration in milliseconds

	‘no barrier’	VOC
9-10 yrs	859 (14)	883 (15)
13-14 yrs	888 (17)	964 (25)
Adults	699 (09)	765 (10)

3.2. Word duration

Repeated-measures ANOVAs were carried out on mean word durations (see Table 2). 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, with all three

groups significantly increasing their mean word duration in VOC. Bonferroni pairwise comparisons indicated that mean word durations were significantly longer for children than both teens and adults, but that the latter two did not differ.

Table 2. Means and standard deviations in parentheses for mean word duration in seconds for the three age groups and both task conditions

	‘no barrier’	VOC
9-10 yrs	.310 (.03)	.420 (.10)
13-14 yrs	.282 (.02)	.351 (.05)
Adults	.262 (.03)	.332 (.06)

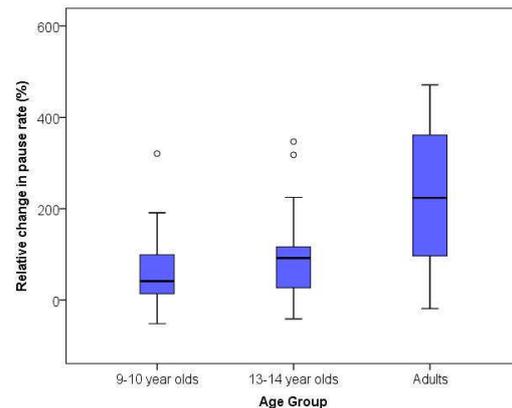


Figure 1: Box-plot showing the measure of relative change in pause rate for talker A across the ‘no barrier’ and VOC conditions. Two outliers in the adult group are not shown.

3.3. Classification of repair strategies

Total counts of direct requests for clarification for each of the three age groups (20 dialogues per group) are given in Table 3. The count of stacked requests, where multiple turns were needed for comprehension to occur, is also given. A univariate ANOVA revealed that there was a significant effect of age group for the total number of direct requests [$F(2,57)=12.03; p<0.0001$]; Bonferroni-adjusted pairwise comparisons showed that there were a greater number of direct requests in child and teen recordings than in adult recordings, but that rates for children and teens did not differ. Group differences were also seen in the number of stacked requests: there were only 4 instances of stacked requests in adult recordings (versus 18 for teens and 37 for children) showing that clarifications made by adults were typically successful at the first turn, and that this was not the case for children or teens.

The types of clarifications made following direct requests were then examined in more detail (see Table 4 and Figure 2). These were calculated, for each dialogue, as a proportion of the total number of requests in that dialogue. One dialogue in the adult group and one in the teen group contained no requests for clarifications so 19 dialogues were included in the analyses for these two groups. The proportion of repetitions varied significantly as a function of age group [$F(2,55)=11.29; p<0.0001$]; adults used significantly less direct repetitions than children, but the difference in repetition rate between teens and the other two groups just failed to reach significance. Data for the other three strategies (expansion, rephrasing, reduction)

violated Levene's test of the equality of variances so non-parametric (Kruskal-Wallis) tests were run. None reached significance although there was a tendency for adults to use more rephrasing strategies.

Table 3. For each age group, total count of direct requests for clarification and of 'stacked' requests, as well as means for both measures per dialogue.

requests	9-10 yrs		13-14 yrs		Adults	
	Sum	Mean	Sum	Mean	Sum	Mean
Direct	222	11.1	174	8.7	77	3.9
Stacked	37	1.9	18	0.9	4	0.2

Table 4. Mean proportion of repetitions, expansions, rephrasing and reduction strategies per age group.

Strategies	9-10 yr olds		13-14 yr olds		Adults	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Repetition	0.51	0.22	0.32	0.26	0.13	0.27
Expansion	0.16	0.16	0.25	0.22	0.24	0.30
Rephrasing	0.15	0.11	0.15	0.18	0.30	0.29
Reduction	0.19	0.14	0.28	0.25	0.28	0.28

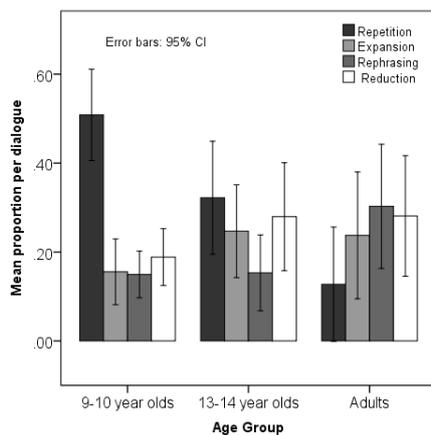


Figure 2: Mean proportion of repetitions, expansions, rephrasing and reduction strategies in response to a direct request for clarification.

4. Discussion

This study examined speech timing strategies (speech rate and pausing) used in communications involving an 'impaired' interlocutor as well as the use of discourse repair strategies following direct requests for clarification. Our prediction was that some of these strategies would be late acquired and that adults would use a broader range of both timing and repair strategies in difficult communications than children, with teens showing greater abilities than children although potentially not reaching adult norms.

Reducing speech rate in the VOC condition was a strategy that was equally available to all three age groups; children spoke at a slower speech rate overall than teens and adults, which reflects previous findings [22]. However, a different picture

was obtained in terms of the use of pauses for clarifying speech. This strategy was used by adults, who showed greater relative change in pause rate across conditions than children or teens. Pauses tended to be of longer duration in the speech of children and teens, which may be a marker of greater cognitive effort involved in discourse planning, but may also be influenced by a slower speech rate.

The rate of direct requests for clarification can inform us on how successfully talkers clarified their speech to counter the effects of adverse listening conditions for their interlocutor, although they are also influenced by the level of perceptual difficulty experienced by interlocutors. The significantly lower rate of direct requests for adults relative to children or teens suggests that adults were more successful at making the acoustic-phonetic and linguistic adaptations to their speech to maintain effective communication. It is notable that 13-14 year olds were still not adult-like in their capacity to clarify their speech. The fact that the same teens did not use vowel hyperarticulation to the same extent as adults in these recordings [15] suggests that they are not using the full range of adaptations typically seen in adult talkers.

Previous studies of repair strategies used by children in stacked requests for clarification [e.g., 18] showed developmental trends, with a greater variety of repair strategies used by 9 year olds than by younger children. Another study with 11-18 year olds found a rate of direct repetitions of 38% which tallies well with our data, although no adult data was given in comparison [23]. Here, children and teens showed a preponderance of straight repetitions of the utterance that had not been understood by their interlocutor, and fewer instances of other repair strategies, while adults used expansion, reduction and rephrasing strategies more often than direct repetitions. The high degree of within-group variability in all groups should be noted and to a certain extent obscures trends such as the reduction in direct repetitions in teens relative to children. This variability may reflect individual differences within each age group, and the use of proportional data for dialogues with few direct requests.

5. Conclusions

To achieve successful communication in a variety of environments, talkers need to continuously adapt acoustic-phonetic and linguistic aspects of their speech production [9]. Our data suggest that developments in speech adaptation strategies are ongoing until mid adolescence at least, although some basic strategies are already established within the first few years of life. Together with findings that perception of noisy and reverberant speech is not adult-like until 14 years [24, 25], this suggests that adverse listening environments are likely to have a deleterious effect on communication for children and teens throughout their school years.

6. Acknowledgements

This work was supported by the Economic and Social Research Council [grant number: RES-062-23-3106]. We gratefully acknowledge Tom Rice-Hird's contribution to the analysis of the dialogue transcripts.

7. References

- [1] Picheny, M. A., Durlach, N. I., and Braidá, L. D., "Speaking clearly for the hard of hearing. II. Acoustic characteristics of clear and conversational speech", *J. Speech Hear. Res.*, 29: 434–446, 1986.
- [2] Smiljanic, R. and Bradlow, A. R., "Production and perception of clear speech in Croatian and English", *J. Acoust. Soc. Am.*, 118 (3): 1677-1688, 2005.
- [3] Uchanski, R. M., Choi, S., Braidá, L. D., Reed, C. M., and Durlach, N. I., "Speaking clearly for the hard of hearing. IV. Further studies of the role of speaking rate", *J. Speech Hear. Res.*, 39: 494–509, 1996.
- [4] Liu, S., and Zeng, F.-G., "Temporal properties in clear speech perception", *J. Acoust. Soc. Am.*, 120: 424–432, 2006.
- [5] Krause, J. C., and Braidá, L. D., "Acoustic properties of naturally produced clear speech at normal speaking rates", *J. Acoust. Soc. Am.*, 115:362–378, 2004.
- [6] Smiljanic, R. and Bradlow, A., "Speaking and hearing clearly: Talker and listener factors in speaking style changes", *Language and Linguistics Compass*, 3 (1): 236-264, 2009.
- [7] Cooke, M., and Lu, Y., "Spectral and temporal changes to speech produced in the presence of energetic and informational maskers," *J. Acoust. Soc. Am.*, 128:2059–2069, 2010.
- [8] Hazan, V. and Baker, R., "Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions", *J. Acoust. Soc. Am.* 130: 2139-2152, 2011.
- [9] Lindblom, B. (1990) "Explaining phonetic variation: a sketch of the H&H theory". in W. J. Hardcastle and A. Marchal [eds] *Speech Production and Speech Modelling*, The Netherlands: Kluwer Academic, pp. 403-439.
- [10] Redford, M.A., Gildersleeve-Neumann, C.E., "The development of distinct speaking styles in preschool children", *J. Speech Lang. Hear. Res.*, 52: 1434-1448, 2009.
- [11] Lotto, A., Ide-Helvie, D.L., McCleary, E.A., Higgins, M.B. "Acoustics of clear speech from children with normal hearing and cochlear implants", *J. Acoust. Soc. Am.*, 119: 3341, 2006.
- [12] Lee, S., Potamianos, A., & Narayanan, S., "Acoustics of children's speech: Developmental changes of temporal and spectral parameters", *J. Acoust. Soc. Am.*, 105: 1455-1468, 1999.
- [13] Gerosa, M., Giuliani, D. and Brugnara, F., "Acoustic variability and automatic recognition of children's speech", *Speech Commun.* 49: 847-860, 2007.
- [14] Baker, R., and Hazan, V., "DiapixUK: a task for the elicitation of spontaneous speech dialogs", *Behav. Res. Meth.*, 43: 761-770, 2011.
- [15] Pettinato, M. and Hazan, V., "The development of clear speech strategies in 9-14 year olds", *Proceedings of Meetings on Acoustics (POMA)*, in press.
- [16] Smiljanic, R., and Bradlow, A.R., "Stability of temporal contrasts across speaking styles in English and Croatian", *J. Phon.*, 36: 91–113, 2008.
- [17] Varonis, E., & Gass, S., "Non-native/non-native conversations: A model for the negotiation of meaning", *Applied Linguistics*, 6: 71-90, 1985.
- [18] Brinton, B., Fujiki, M., Frome Loeb, D., & Winkler, E., "The development of conversational repair strategies in response to requests for clarification", *J. Speech Hear. Res.*, 29: 75-81, 1986.
- [19] Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M. and Bradlow, A. R., "The Wildcat Corpus of Native- and Foreign-Accented English: Communicative efficiency across conversational dyads with varying language alignment profiles", *Lang. Speech*, 53: 510-540, 2010.
- [20] Boersma, P. & Weenink, D., "Praat: doing phonetics by computer" [Computer program]. Version 5.3.07, <http://www.praat.org/>. (last retrieved 5 March 2012.)
- [21] Tang, Y., Nicolao, M., Raitio, T., Jokinen, E., Koutsogiannaki, M., Godoy, E., Granlund, S., Aubanel, V., Sfakianaki, A., Mowlae, P., Hazan, V. & Stylianou, Y. "exPlic8 acoustic-phonetic analysis software" [Computer program]. Version 1.0. (last retrieved 10 January 2013).
- [22] Amir, O. and Grinfeld, D., "Articulation Rate in Childhood and Adolescence: Hebrew Speakers", *Language and Speech*, 54: 225-240, 2011.
- [23] Most, T., "The use of repair strategies by children with and without hearing impairment. *Language*", *Speech, and Hearing Services in Schools*, 33(2):112–123, 2002.
- [24] Elliott L.L., "Performance of children aged 9–17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability", *J. Acoust. Soc. Am.*, 66: 651-653, 1979.
- [25] Johnson CE., "Children's phoneme identification in reverberation and noise", *J Speech Lang Hear Res.*, 43:144-157, 2000.