DO CHILDREN ENHANCE PHONETIC CONTRASTS IN SPEECH DIRECTED TO A HEARING-IMPAIRED PEER?

Sonia Granlund[1], Valerie Hazan[1] and Merle Mahon[2]

[1]Speech Hearing and Phonetic Sciences; [2]Developmental Science;
UCL (University College London)
s.granlund@ucl.ac.uk

ABSTRACT

This study examines whether normal-hearing (NH) children enhance phonetic contrasts when speaking to a hearing-impaired (HI) peer. A problem-solving 'Grid' task was developed to elicit frequent repetitions of /p/-/b/, /s/-/ʃ/ and /i/-/ɪ/ segmental contrasts and point vowels in communicative spontaneous speech. Eighteen NH children between 9 and 15 years old performed the task once with a NH friend and once with a HI friend. Both category means and within-speaker variability were analysed. Results suggest that although HI interlocutors are likely to find the phonetic contrasts difficult to produce and perceive, children's HI-directed speech contains little evidence of phonetic category enhancement.

Keywords: clear speech, intra-speaker variability, speaker-listener adaptations, children's speech

1. INTRODUCTION

Clear speech research aims to describe the acoustic-phonetic modifications to speech made by speakers talking to a listener in adverse listening conditions. Studies propose that 'global' clear speech modifications, such as increased f0 range and intensity, are accompanied by a speaker's attempt at enhancing phonological contrasts to listeners with a greater approximation of phonetic targets [3, 14]. Most evidence for this hypothesis comes from research examining clear speech vowel space size, which is found to be increased in clear speech compared to 'casual' speech [2, 5, 14]. However, speakers of languages with different sizes of vowel inventories have been found to increase vowel space size to similar degrees in clear speech [25, 2], casting doubt on whether vowel space is specifically enhanced to make vowels more contrastive from each other. Few studies have examined other segmental contrasts, with findings suggesting that speakers enhance stop voicing distinctions according to the speaker's primary cue [15], and that the fricative spectral distinction /s/-/ʃ/ is made more contrastive in clear speech [18]. On the other hand, speakers do not enhance tense-lax vowel distinctions according to the primary vowel cue in the language [9], and voice-onset-time (VOT) distinctions were not found to change for clear speech stops [26].

Enhancing phonetic contrasts by approximating phonetic targets more closely may indeed be helpful to listeners – there is evidence that speakers who have more separable and internally more consistent phonetic categories, i.e. greater distances between phonemes, less within-category dispersion and less overlap between categories, are more intelligible to listeners than those whose categories are less discriminable [12, 20]. However, clear speech studies typically only report on mean measures per category without examining discriminability and variability within categories.

An important limitation of previous studies is also the use of read sentences and instructions, such as 'speak as if to a hearing-impaired person' [3, 26] when eliciting clear speech. There is evidence that different instructions affect clear speech production [16, 24] and read clear speech may differ from spontaneously induced clear speech [10]; in natural speaking situations, speakers may hyperarticulate based on listener feedback [18].

Additionally, most clear speech studies investigate adult speakers. Very little is known on whether children are able to make speech adaptations according to the needs of their listener. Preschoolers have been found to enlarge their vowel spaces when teaching a toy puppet to speak [27], but their vowel formant frequencies did not change when asked to speak clearly 'in a big girl/boy voice' [22]. Initial evidence from older school-aged children suggests that they may not be able to make as specific spontaneous global clear speech adjustments as adults do [13], perhaps due to their still-developing speech motor system. None of the above studies have specifically investigated clear speech phonetic contrast enhancement in older children. Indeed, children may be more variable in their speech production even until the age of 14 [17], which may make segmental contrast enhancement more difficult. It is especially important to investigate whether school-aged normally-hearing (NH) children are able to modify their speech to the needs of a hearing-impaired (HI) listener as, in the UK, over 80% of HI children attend mainstream schools with NH peers [4].

This study aims to address these shortcomings by exploring whether older NH children enhance the discriminability of phonetic contrasts when in spontaneous task-based conversation with a HI friend compared to a NH friend. We use a novel problem-solving task to elicit spontaneous production of phonetic contrasts which are likely to be difficult for the HI children to both produce and perceive. Clarifying these contrasts in the task had a specific communicative role, as listeners had to perceive the keywords accurately to be able to complete the task. The interlocutor may therefore need to enhance these contrasts in HI-directed speech, but not in NH-directed speech. As well as measuring vowel space area and category means for several different types of phonetic contrasts, we also measure category dispersion and overlap to obtain a more comprehensive assessment of segmental contrast enhancement by children in a realistic speaking situation. If speakers attempt to approximate phonetic targets when speaking clearly, we are likely to find greater between-category distances, less within-category dispersion and less overlap between categories in HI-directed speech compared to NH-directed speech.

2. METHOD

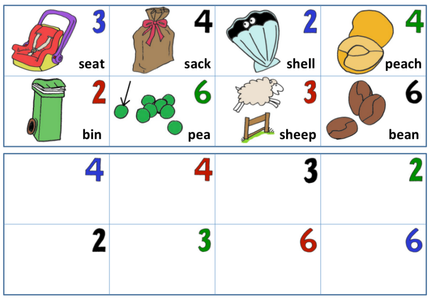
2.1. Participants

Eighteen NH children (mean age: 11.9 years; range: 9.0-14.3; 11 female) from Southern England participated in the experiment[1]. Another eighteen HI children (mean age: 12.0 years; range: 9.7-15.2; 10 female) acted as confederates in the HI-directed condition. All participants were students at mainstream primary or secondary schools with an adjacent unit for hearing-impaired students. Four friends (2 NH; 2 HI) were recruited per school. Of the NH participants, three were bilingual, and another two had mild additional needs.The HI participants' hearing loss levels ranged from moderate (4 HIs) to profound (7 HIs); eleven participants wore digital hearing aids, and the remaining seven participants wore at least one cochlear implant (CI).

2.2. Materials and Procedure

The 'Grid' task was developed to enable the elicitation of several repetitions of keywords containing different types of segmental contrasts in communicative spontaneous speech. The bilabial voicing contrast /p/-/b/, the sibilant place distinction /s/-/ʃ/ and the high vowel contrast /i/-/ɪ/ were chosen as they have typically been found to be difficult for HI children to produce and perceive [19, 8]; therefore the contrasts may need to be enhanced in HI-directed speech. These distinctions also represent different kinds of phonetic contrasts (temporal, spectral and spectro-temporal) which may be enhanced to different degrees. Sixteen minimal pair keywords containing these contrasts were created, three per contrast[2]. Five versions of each keyword were hand-drawn (e.g., five different kinds of sheep). The keywords were randomised, and a version of each keyword was placed on eight different four-by-two picture-grids. Each grid square also contained a coloured number; the colours and numbers were chosen to elicit as wide a range of vowels as possible. An empty four-by-two grid with only coloured numbers was placed underneath each picture-grid (see Fig.1).

Figure 1: An example picture-grid and empty grid given to each participant in the task.



The pair of participants sat opposite each other; each participant was given a different version of the picture-grid, an empty grid, and a tray of cards containing the five different representations of each keyword. The participants could see each other but not each others' grids. The aim of the task was for the pair to build up each others' picture-grids in their empty grids. For each square on their picture-grid, each speaker had to describe to the other: i) the keyword; ii) the picture version of the keyword; and iii) the 'location' (coloured number) of the picture on their grid. The interlocutor had to find in her card tray the correct picture representation of the correct keyword and attach it to the correct location on her empty grid.

Within the group of four from each school, each participant completed two 'communication' sessions: one with a NH friend (to elicit NH-directed speech), and the other with a HI friend (to elicit

HI-directed speech). Half of the participants completed the first session with a NH friend. Note that although each HI participant also completed the task with both a NH and HI friend, the HI participant's speech is not analysed in this study; the task was part of a wider experimental protocol not reported here. Each pair of participants completed between two and four grids together per session, with each session lasting between 1 and 1.5 hours. All sessions were recorded at the children's schools in a quiet room. Both participants wore a high-quality lapel microphone and were audio- and video-recorded.

2.3. Processing

Each participant's speech was saved to a separate audio channel; in total, 36.3 hours of single-channel audio recordings were made. They were transcribed, and the transcriptions auto-aligned to the waveform on word and segmental level. The resulting Praat [1] TextGrids were hand-checked on the word level. For all measures, the parts of the signal containing the interlocutor's speech were excluded from analysis. Unintelligible words, words spoken while laughing, partially spoken words and silences were not analysed.

While transcribing the keywords, the transcriber ensured that the transcribed keyword was the one intended to be produced by the speaker. The TextGrid files were used to find all instances of the keywords in the Grid task, and all target sounds in each of the keywords were then manually segmented using Praat. After segmentation, Praat scripts were used to measure the data; the VOT of /p/ and /b/ and the first spectral moment (centre of gravity) for the mid 50% of /s/-/ʃ/ segments using DFT spectra was calculated. The duration of each /i/-/ɪ/ segment was also measured, and the midpoint F1 and F2 of each vowel was determined. F1 and F2 outliers were excluded. Formant values were then normalised to ERB and, to obtain one spectral measure per vowel, the Euclidean distance between F1 and F2 was taken for each vowel, calculated as $\sqrt{(F1-F2)^2}$. Altogether, 2554 segments extracted from the keywords were included in the analyses. On average, 12.5 tokens of each phoneme were analysed per speaker per condition. The measure of between-category distance was taken as the difference in the mean of phoneme 1 and the mean of phoneme 2. Within-category dispersion was calculated as the mean of the standard deviations for the two phonemes in the contrast, and category overlap was the minimum of phoneme 1 subtracted from the maximum of phoneme 2 ([23]), per speaker per condition.

To obtain a measure of vowel space area, the vowel midpoint F1 and F2 of /i/, /æ/ and /ɔ/ in all content words in the Grid task were measured using a Praat script. Outliers were removed, and a subset of values were manually checked. Formant values were normalised to ERB. Vowel space area (one value per speaker per condition) was then calculated on the basis of 4240 vowels, with approximately 41.6 tokens of each vowel analysed per speaker per condition[3].

3. RESULTS

3.1. Task difficulty

First, to explore whether participants found the Grid task more difficult in conditions involving a HI interlocutor, Wilcoxon rank sum tests were carried out on task transaction times. Results show that NH-NH pairs took less time to find each picture-square correctly in the Grid task (mean: 25.0s) than did NH-HI pairs (mean: 45.5s) (W=128, p=0.004). Because of the increased difficulty of completing the task with their HI friend, participants may need to enhance phonetic contrasts in that condition.

3.2. Phonetic contrast enhancement

To investigate whether speakers make phonetic contrasts more distinct in their speech when talking with a HI interlocutor compared to a NH interlocutor, the three phonetic contrasts were explored in terms of their category distinctiveness in HI-directed versus NH-directed conditions. For each contrast, the Shapiro-Wilk command in R [21] was used to test each measure for normality. For normally distributed data, paired t-tests were performed; for non-normally distributed data, Wilcoxon signed rank tests were used. Effect size r was calculated following [7]. Raw values for each phonetic contrast can be seen in Table 1.

3.2.1. Temporal /p/-/b/ contrast

For the /p/-/b/ contrast, measured in VOT, there were no significant differences between NH- and HI-directed speech for any of the between-category distance (p=0.27), within-category dispersion (p=0.15) or category overlap (p=0.19) measures. This implies that NH speakers do not enhance the VOT distinction – either by making category distances greater, category dispersion narrower or by reducing category overlap – when talking with a HI friend.

3.2.2. Spectral /s/-/ʃ/ contrast

Similarly, for the /s/-/ʃ/ contrast, measured in spectral centre of gravity, there were no significant dif-

Table 1: Means (with SDs in parentheses) for the phonetic contrasts, in VOT (/p/-/b/), centre of gravity (/s/-/ʃ/), Euclidean distance and duration (/i/-/ɪ/), in NH-directed (NHD) and HI-directed (HID) conditions.

phoneme	NHD	HID
/p/ (ms)	78 (33)	77 (31)
/b/ (ms)	8 (2)	0 (29)
/s/ (Hz)	7762 (1784)	7987 (1382)
/ʃ/ (Hz)	4531 (804)	4725 (762)
/i/ (ERB)	14.4 (1.0)	14.2 (1.2)
/ɪ/ (ERB)	10.9 (1.2)	11.4 (1.3)
/i/ (ms)	196 (82)	218 (105)
/ɪ/ (ms)	134 (47)	139 (43)

ferences between HI- and NH-directed speech for any of the measures (distance: p=0.83; dispersion: p=0.22; overlap: p=0.53). These results suggest that the speakers do not enhance the contrast in HI-directed speech.

3.2.3. Spectro-temporal /i/-/ɪ/ contrast

For the spectral primary cue to the /i/-/ɪ/ contrast, measured in Euclidean distance in ERB, the difference between HI-directed and NH-directed speech was significant for all measures. However, all effects were in the opposite direction to that expected: speakers had *smaller* overall distances between /i/-/ɪ/ categories in HI-directed (mean: 3.0) than in NH-directed (mean: 3.5) speech (t(16)= -3.7, p=0.002, r=0.68); within-category dispersion was *greater* in HI-directed (mean: 1.0) than in NH-directed (mean: 0.82) speech (t(16)=2.2, p=0.04, r=0.48); and /i/-/ɪ/ category overlap was greater in HI-directed (mean: -0.67) than in NH-directed (mean: 0.78) speech (t(16)=-2.55, p=0.01, r=0.58). Therefore, spectrally, speakers produced less discriminable /i/-/ɪ/ spectral contrasts in HI-directed speech than in NH-directed speech.

For the secondary cue to the /i/-/ɪ/ contrast – duration – there were no significant differences between HI- and NH-directed speech for any of the measures (distance: p=0.26; dispersion: p=0.26; overlap: p=0.50). Therefore, speakers were not found to enhance the temporal distinction between /i/ and /ɪ/ in HI-directed speech.

3.3. Vowel space area enhancement

To examine whether speakers enlarge the size of their vowel space in HI-directed speech, vowel space area (in ERB^2) was analysed with the same statistical method as above, but using only mean vowel space area as a measure per speaker per condition. Vowel space area was found to be enhanced in HI-directed speech; the area was slightly larger in HI-directed (mean: 22.3) than in NH-directed (mean: 20.9) speech (t(16)=2.04, p=0.058, r=0.45).

3.4. Effect of gender

To ensure that the above results did not occur due to female and male participants being analysed together, a mixed ANOVA was run on speaker gender and 'directed' (NH-directed; HI-directed) on each of the measures. There were no significant interactions ($0.97 \geq p \geq 0.31$), suggesting that speaker gender did not influence the results.

4. DISCUSSION

In summary, our findings suggest that, despite the increased difficulty in completing the task with a HI friend, most phonetic category distinctions were not enhanced in HI-directed speech. Speakers did not increase the VOT distinctions between /p/ and /b/ categories, nor the spectral centre of gravity between /s/ and /ʃ/. Surprisingly, the secondary cue for the /i/-/ɪ/ contrast, duration, was also not enhanced in HI-directed speech, despite some adult clear speech studies' findings to the contrary [9].

Although, as in [2, 5, 14], a tendency towards vowel space enhancement was found, spectral tense-lax distinctions for /i/-/ɪ/ were *reduced* (as in [9]). This suggests that spectral changes in vowels in clear speech may in fact be a side-effect of other clear speech changes, such as increased vocal intensity [6], rather than a feature of phonological contrast enhancement.

The participants in the current study were children, and therefore the lack of contrast enhancement could be attributed to their still-developing speech production system [17, 13]. Indeed, children as old as 12 years have been found to categorise phonemic contrasts less consistently than adults [11] – therefore even school-aged children may not have accurate knowledge of the cues that would be useful in making their phonetic contrasts more discriminable. However, it is striking that, in the current study, phonetic contrasts were not enhanced even in the Grid task, in which a speaker's increased category discriminability would have been very useful for completing the task successfully. It is possible that, rather than using acoustic-phonetic modifications, children use other, perhaps easier, strategies available to them, such as increasing linguistic contextual information to aid communication with their HI interlocutor.

5. ACKNOWLEDGMENTS

We thank all the participating schools, parents and children. This project was funded by a PhD studentship from the UK Economic and Social Research Council (ESRC) linked to grant number RES-062-23-3106.

6. REFERENCES

[1] Boersma, P. and Weenink, D., 2014. *Praat: doing phonetics by computer, version 5.3.82*.

[2] Bradlow, A. R. 2002. Confluent talker-and listener-oriented forces in clear speech production. *Lab. Phon.* 7, 241–273.

[3] Bradlow, A. R., Bent, T. 2002. The clear speech effect for non-native listeners. *J. Acoust. Soc. Am.* 112(1), 272.

[4] CRIDE, 2014. Consortium for Research in Deaf Education: Report on 2014 survey on educational provision for deaf children in England; http://www.ndcs.org.uk/document.rm?id=9796.

[5] Ferguson, S. H., Kewley-Port, D. 2007. Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *J. Sp. Lang. Hear. Res.* 50(5), 1241–1255.

[6] Ferguson, S. H., Quené, H. 2014. Acoustic correlates of vowel intelligibility in clear and conversational speech for young normal-hearing and elderly hearing-impaired listenersa). *J. Acoust. Soc. Am.* 135(6), 3570–3584.

[7] Field, A., Miles, J., Field, Z. 2012. *Discovering Statistics Using R*. London: Sage.

[8] Giezen, M. R., Escudero, P., Baker, A. 2010. Use of acoustic cues by children with cochlear implants. *J. Sp. Lang. Hear. Res.* 53(6), 1440–1457.

[9] Granlund, S., Hazan, V., Baker, R. 2012. An acoustic–phonetic comparison of the clear speaking styles of Finnish–English late bilinguals. *Journal of Phonetics* 40(3), 509–520.

[10] Hazan, V., Baker, R. 2011. Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *J. Acoust. Soc. Am.* 130(4), 2139.

[11] Hazan, V., Barrett, S. 2000. The development of phonemic categorization in children aged 6–12. *Journal of Phonetics* 28(4), 377–396.

[12] Hazan, V., Romeo, R., Pettinato, M. 2013. The impact of variation in phoneme category structure on consonant intelligibility. *ICA 2013 Montreal*. Proc. Meetings on Acoustics 1–6.

[13] 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. *J. Sp. Lang. Hear. Res.*

[14] Johnson, K., Flemming, E., Wright, R. 1993. The hyperspace effect: Phonetic targets are hyperarticulated. *Language* 505–528.

[15] Kang, K.-H., Guion, S. G. 2008. Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. *J. Acoust. Soc. Am.* 124(6), 3909–3917.

[16] Lam, J., Tjaden, K., Wilding, G. 2012. Acoustics of Clear Speech: Effect of Instruction. *J. Sp. Lang. Hear. Res.* 55(6), 1807.

[17] Lee, S., Potamianos, A., Narayanan, S. 1999. Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *J. Acoust. Soc. Am.* 105(3), 1455–1468.

[18] Maniwa, K., Jongman, A., Wade, T. 2009. Acoustic characteristics of clearly spoken English fricatives. *J. Acoust. Soc. Am.* 125(6), 3962.

[19] Mildner, V., Šindija, B., Vrban Zrinski, K. 2006. Speech perception of children with cochlear implants and children with traditional hearing aids. *Clinical Linguistics & Phonetics* 20(2-3), 219–229.

[20] Newman, R., Clouse, S., Burnham, J. 2001. The perceptual consequences of within-talker variability in fricative production. *J. Acoust. Soc. Am.* 109, 1181–1196.

[21] R Core Team, 2014. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria.

[22] Redford, M., Gildersleeve-Neumann, C. 2009. The development of distinct speaking styles in preschool children. *J. Sp. Lang. Hear. Res.* 52, 1434–1448.

[23] Romeo, R., Hazan, V., Pettinato, M. 2013. Developmental and gender-related trends of intra-talker variability in consonant production. *J. Acoust. Soc. Am.* 134(5), 3781.

[24] Scarborough, R., Brenier, J., Zhao, Y., Hall-Lew, L., Dmitrieva, O. 2007. An acoustic study of real and imagined foreigner-directed speech. *Proc. 16th ICPhS* 2165–2168.

[25] Smiljanić, R., Bradlow, A. R. 2005. Production and perception of clear speech in Croatian and English. *J. Acoust. Soc. Am.* 118(3), 1677.

[26] Smiljanić, R., Bradlow, A. R. 2008. Stability of temporal contrasts across speaking styles in English and Croatian. *Journal of Phonetics* 36(1), 91–113.

[27] Syrett, K., Kawahara, S. 2014. Production and perception of listener-oriented clear speech in child language. *Journal of Child Language* 41(6), 1373–1389.

---

[1] One female participant was excluded due to equipment malfunction in one recording session.

[2] pin-bin, peach-beach, pea-bee; cell-shell, seat-sheet, sack-shack; bean-bin, peach-pitch, sheep-ship.

[3] Mean distances between /i-æ/ (a), /i-ɔ/ (b) and /ɔ-æ/ (c) were calculated, and Heron's method was used to calculate vowel space: $\sqrt{s*((s-a)*(s-b)*(s-c))}$; where s=(a)+(b)+(c).