HOW DOES FOREIGNER-DIRECTED SPEECH DIFFER FROM OTHER FORMS OF LISTENER-DIRECTED CLEAR SPEAKING STYLES?

Valerie Hazan[1], Maria Uther[2], Sonia Granlund[1]

[1]Speech Hearing and Phonetic Sciences, UCL (University College London), UK; [2] Department of Psychology, University of Winchester, UK
v.hazan@ucl.ac.uk, Maria.Uther@winchester.ac.uk, s.granlund@ucl.ac.uk

ABSTRACT

Forty talkers participated in problem-solving tasks with another talker in conditions differing in communication difficulty for the interlocutor. A linguistic barrier condition (L2 interlocutor) was compared to acoustic barrier conditions (native interlocutors hearing vocoded or noisy speech). Talkers made acoustic-phonetic enhancements in all barrier conditions compared to the no-barrier condition, but talkers reduced their articulation rate less and showed less increase in vowel hyperarticulation in foreigner-directed speech than in the acoustic barrier condition, even though communicative difficulty was greater in the L2 condition. Foreigner-directed speech was also perceived as less clear. This suggests that acoustic enhancements in clear speech are not simply a function of the level of communication difficulty.

Keywords: Speech production, speaking styles, foreigner-directed speech.

1. INTRODUCTION

This study is concerned with the acoustic-phonetic adaptations that talkers make in communicative situations to counter the effects of three different types of adverse listening conditions. To produce clear speech, talkers make adaptations to their speech both at a global and segmental level (see [4, 13] for a review). The contrast between 'clear speech' and casual or conversational speech can be interpreted with the Hyper-Hypo (H& H) theory of speech production [9]. According to the H&H model, speech production is goal-directed and aimed at maximising communication: talkers have to dynamically adjust their speech production along a hyper- to hypospeech continuum, as communicative demands change, to maintain communication at the least cost in terms of articulatory effort. In most studies of 'clear speech', this speaking style has primarily been elicited by giving participants instruction to speak 'as if talking to a person who has a hearing impairment or to a non-native speaker'. Although researchers have considered these two types of instruction as eliciting a single category of 'clear speech', it has been shown that different clear speech instructions to talkers producing read materials lead to different magnitudes of speech adaptations [8]. This is not surprising as speech directed at hearing-impaired listeners is mainly aimed at overcoming an acoustic barrier to communication, whilst speech directed at non-native speakers is aimed at overcoming a linguistic barrier. Speech modifications aimed at overcoming acoustic barriers may not benefit listeners facing linguistic barriers, e.g. [3].

In order to better understand the impact of talker-listener interaction on speech production in adverse conditions, two issues need addressing. First, a clearer separation of different kinds of clear speech is needed. Second, analyses should be of recordings where talkers are engaged in more ecologically valid forms of communication, rather than simply being instructed to imagine the interaction, as there are differences in acoustic characteristics between instructed and naturally-elicited speech [11,12]. Addressing both of these issues, a recent study [7] investigated the acoustic-phonetic characteristics of speech directed at interlocutors facing two different types of acoustic barrier. Talkers were engaged in a problem-solving task with an interlocutor who heard them either normally, through a three-channel vocoder (VOC) or in simultaneous babble noise (BAB). The VOC and BAB conditions elicited different speech adaptations in the talker countering the effects of these two acoustic barriers. Here, those findings are extended by considering a third condition, not reported in [7], where talkers interacted with an interlocutor facing a linguistic barrier: a low-proficiency non-native speaker of English (L2 condition). We also examine the impact of these three types of modifications on the perceived clarity of the resulting speech.

This study had two main objectives. The first was to determine whether speech adaptations varied across the linguistic- and acoustic-barrier conditions. We predicted that greater acoustic-phonetic adaptations would be made where clear speech was countering the effects of an acoustic barrier (VOC and BAB) than in foreigner-directed speech (L2). The second was to determine whether speech countering an acoustic barrier was perceived as clearer than speech countering a linguistic barrier. We predicted that speech adaptations made to counter acoustic barriers would result in greater clarity due to the greater degree of enhancement.

2. METHOD

2.1. Speech Corpus

Recordings were taken from the LUCID corpus [1]. This corpus includes a range of recorded materials from 40 native talkers of Southern British English (20 M; 20 F) aged between 19 and 29 years. All participants were screened for normal hearing thresholds and reported no history of speech or language disorders. Another eight native talkers of Southern British English (4 M; 4 F) fulfilling the above criteria were recruited as confederates for the BAB condition. Six non-native talkers (two each from China, Taiwan and Korea) were recruited as confederates for the L2 condition. All non-native confederates scored within the 4th and 5th ability groups for oral proficiency ('basic user' level) on a standardized test of English language skills (VersantTM).

Recordings were made of pairs of talkers carrying out the diapix task [15], which required them to find 12 differences between two cartoon-like pictures without sight of the other talker's picture. The diapixUK set of picture pairs was used; see [1] for a full description of the task and of the picture materials and [7] for a full description of recording conditions. In the 'no barrier' (NB) condition, the two talkers heard each other without any interference. In the VOC condition, the same two talkers carried out the task but Talker B heard Talker A via a three-channel noise-excited vocoder. Both talkers had received around 10 minutes of training in listening to vocoded speech. In the BAB condition, Talker A carried out the task with a confederate hearing in a background of multi-talker babble. In the L2 condition, Talker A carried out the task with a low-proficiency non-native confederate. Three diapix tasks were carried out in each condition. The 40 talkers acting as Talker A were split into two groups: 20 talkers formed the 'VOC/L2' group, recorded in an acoustic (VOC) and a linguistic (L2) barrier condition as well as the control NB condition; the remaining 20 talkers formed the 'VOC/BAB' group, recorded in two different acoustic barrier conditions (VOC, BAB) as well as NB. In the barrier conditions, Talker A was told to take the lead in the interaction. On average, recordings for the barrier conditions lasted 28 minutes, giving around 12 minutes of speech for talker A to be analysed [7].

2.2. Perceived clarity rating experiment

The aim of the clarity rating study was to establish whether there were differences in perceived clarity between speech aimed at overcoming acoustic barriers (VOC, BAB) and a linguistic barrier (L2). Clarity ratings were used because direct measures of intelligibility are difficult to obtain for spontaneous speech due to ceiling effects. Perceived clarity ratings for sentence materials have been shown to be correlated with intelligibility measures [6]. For each talker, nine short speech samples were extracted from the diapix recordings, i.e. 3 samples per talker per condition. Selection criteria were that they be as close to the 20th turn of the interaction as possible, 2-3 seconds long and reasonably self-contained utterances, not preceded by a miscomprehension. There were 120 samples each for NB and VOC, and 60 each for BAB and L2 (Total: 360 samples).

Twenty-four listeners (22 F; mean age: 21.8 years, range: 18.9–28.4) were paid for their participation. All were monolingual Southern British English talkers, with no known speech and language disorders, and hearing thresholds of 20dBHL or better up to 4 kHz. The samples were presented using the ExperimentMFC function in Praat [2] at a comfortable loudness level (68-73dB SPL) through headphones in an acoustically treated room. At the start of the session, participants were told that they would be rating the excerpts for clarity, "i.e. how easy it would be to understand the snippet in a noisy background". After hearing each sample, the listener saw a box asking 'How clear is the speech', with a rating scale from 1(very) to 7(not very) on the screen. The order of the stimuli was randomised and each token was heard once. Participants could take breaks after every 40 presentations if they wished; the test took approximately 30 minutes.

3. RESULTS

3.1. Communication difficulty

A measure of communication difficulty is needed to ascertain that increases in acoustic-phonetic adaptations and in perceived clarity are not simply due to an increase in difficulty across barrier conditions. The number of words produced by Talker A to complete the task (TotalW) in communication barrier conditions was taken as a measure of communication difficulty, as frequent repetitions due to miscomprehensions by Talker B lead to an increase in the number of words produced. This measure also has the advantage over a task-duration measures of being independent of changes in speech rate across conditions. Data were analysed using repeated-measures ANOVAs. For the VOC/L2 group, TotalW was significantly higher for the L2 (*M*=3170, *SD*=699) than for the VOC (*M*=2365, *SD*=579) or NB conditions (*M*=1825, *SD*=390), $F(2,38)=45.2$; $p<.0001$). For the VOC/BAB group, although the effect of condition was just significant [$F(2,38)=3.97$; $p<.05$], with an effect at $p<.06$ found for NB (*M*=1851, *SD*=535) compared to other conditions, TotalW did not differ significantly across VOC (*M*=2191, *SD*=826) and BAB (*M*=2157, *SD*=521).

In summary, the two acoustic barrier conditions were of similar communicative difficulty, as judged by TotalW, while the linguistic barrier led to greater difficulty in completing the task. This was expected as, whilst the acoustic barriers only affect Talker B, the linguistic barrier entails communicative difficulties for both talkers, as both talkers may have difficulty understanding each other's accents.

3.2. Perceived clarity rating

Perceived clarity ratings (see Table 1) were analysed using general linear mixed-effect models using the lme function in the nlme package for R [10]. The best-fitting model was chosen with a hierarchical approach, adding predictors one by one to a baseline model that includes no predictors other than the intercept. Condition (NB, VOC, BAB, L2) and talker sex were entered as fixed effects, and listener (rater) and talker were random effects in all analyses. Likelihood ratio tests were used to determine whether the effect of conditions and talker sex were needed in the model. Tukey's HSD was used for any subsequent post hoc comparisons. The linear mixed effects model's type III Wald Chi square test revealed significant main effects of condition ($\chi2(3) = 669.6$, $p<.0001$) and talker sex ($\chi2(4) = 203.0$, $p<.0001$). The condition*talker sex interaction was significant ($\chi2(7) = 8.1$, $p=0.04$) but was not explored further. Post-hoc analyses show that samples were rated as less clear in NB than VOC ($p< 0.001$), BAB ($p< 0.001$), and L2 ($p< 0.01$). VOC and BAB samples were rated as clearer than L2 samples (both $p<0.001$) but there was no significant difference between VOC and BAB samples ($p=0.986$). Note that this was the case even though rating task instructions could have favoured BAB samples. Women were rated as clearer than men.

In summary, speech produced by Talker A in all conditions in which the interlocutor was affected by a communication barrier were rated as clearer than when there was no communication barrier. Further, speech produced to counter the effects of acoustic barriers to communication was perceived as clearer than speech produced to counter the effects of a linguistic barrier (foreigner-directed speech). This suggests that, although the task was more difficult for Talker A in the L2 condition than in the VOC condition, as shown by greater communication difficulty, the acoustic-phonetic aspects of the speech were not more greatly enhanced in this condition. This is checked further by carrying out acoustic analyses.

Table 1: clarity ratings on a scale of 1 (clear) to 7 (not very clear); standard deviations are given in parentheses.

Talker Group	NB	L2	VOC	BAB
VOC/BAB (N=20)	3.00 *(.46)*	-	2.46 *(.47)*	2.52 *(.47)*
VOC/L2 (N=20)	3.44 *(.64)*	3.10 *(.60)*	2.49 *(.48)*	-

3.3. Acoustic-phonetic analyses

Acoustic-phonetic analyses for the NB, VOC and BAB recordings can be found in [7]. Here, these same measures are presented for the L2 condition to establish which acoustic-phonetic features were adapted when speaking to a non-native interlocutor. As participants acting as Talker A in the L2 condition also carried out the task in VOC, an acoustic barrier condition, the two types of adaptations are compared. Analysis methods are briefly summarised here; for further details, see [7]. Each channel of the dialogue was orthographically transcribed and aligned to the speech signal at word and phoneme levels. All analyses were carried out using Praat scripts on recordings for three diapix tasks per condition. Measures of F0 median and range were obtained. Mean word duration was calculated by dividing the total duration of speech regions by the number of words produced. Median vowel formant values were measured for all monophthongs uttered in content words in each condition and converted to ERB, an auditory scale. Vowel F1 and F2 range values were based, for each talker, on the difference between lowest and highest median F1 values and lowest and highest median F2 values across the talker's vowel range. For intensity, a long-term average spectrum was calculated over all speech regions and a 1-3 kHz mean energy value (ME13) obtained as the mean between these frequencies (See Table 2).

Measures of percentage change relative to the values in the NB condition are used to normalise for individual differences in their NB speech (see Table 3). In L2, relative to NB, talkers slowed down their speech, increased their vowel F1 and F2 ranges, showed a mild increase in intensity in the mid-frequency region and minimal changes in median F0 and F0 range. The relative measures of adaptations were compared for the L2 and VOC conditions in the same group of 20 talkers (see Table 3 for paired t-tests). Mean word duration, vowel F1 and F2 ranges increased significantly less in L2 than VOC, while similar (and small) increases across conditions were obtained for mean energy, median F0 and F0 range. Note it was argued in [7] that changes in intensity and in F0 characteristics would not be helpful in countering the effects of hearing vocoded speech so this might explain the lack of difference between conditions for these measures.

Table 2: Acoustic measures for 20 talkers in the VOC/L2 group. For each recording condition (NB, VOC, L2), the means are calculated across all 20 talkers (standard deviations in parentheses).

	NB	VOC	L2
Mean word duration (sec)	0.25 *(.02)*	0.32 *(.04)*	0.30 *(.03)*
Mean energy 1-3 kHz (dB)	24.8 *(2.2)*	26.5 *(2.5)*	26.8 *(2.3)*
Median F0 (semitone)	86.2 *(6.2)*	87.3 *(6.1)*	87.3 *(5.9)*
F0 range (semitone)	3.1 *(0.8)*	3.2 *(0.5)*	2.9 *(0.5)*
Vowel F1 range (ERB)	3.6 *(1.0)*	4.5 *(1.0)*	4.1 *(0.9)*
Vowel F2 range (ERB)	4.3 *(1.1)*	5.5 *(0.8)*	5.1 *(0.9)*

Table 3: Mean percent change relative to NB condition for the six acoustic phonetic measures in the L2 and VOC conditions. Base units: milliseconds (mean duration), dB (mean energy), semitones relative to 1 Hz (F0) and ERB (vowel ranges). Values for the paired t-tests comparing the two conditions are given in the final two columns.

% change rel.NB	L2	VOC	t(19)	p
word duration	20.6	28.4	3.125	0.006
ME1-3 kHz	8.3	6.9	-1.059	0.303
Median F0	1.3	1.2	-0.091	0.929
F0 range	-0.5	8.0	1.904	0.072
F1 range	19.1	30.6	2.092	0.050
F2 range	23.5	34.7	2.991	0.008

A composite measure was calculated as the mean relative change, per talker, over the six acoustic-phonetic measures. For VOC, a common condition to both talker groups, very similar means were obtained across groups (*M*=18.3 for the VOC/L2 group, *M*=18.4 for the VOC/BAB group), showing the stability of this measure. For the VOC/BAB group, greater changes were made in BAB (*M*=25.0, *SD*=8.9) than in VOC (*M*=18.4, *SD*=10.3); t(19) =3.476 , p= 0.003). For the VOC/L2 group, less enhancement overall was made in the foreigner-directed speech (*M*=12.2, *SD*=10.5) than in VOC (*M*=18.3, *SD*=11.8), t(19) =2.753 , p= 0.013.

4. DISCUSSION

In summary, relative to when communication was unimpaired, talkers enhanced acoustic-phonetic aspects of their speech when engaged in interactions with interlocutors whose perception was impaired either by an acoustic or linguistic barrier, but the type and extent of acoustic-phonetic enhancement that was made was dependent on the barrier.

Foreigner-directed speech (FDS) was characterised by increases in mean word duration and vowel hyper-articulation, as also found in [11, 14]. Vowel hyperarticulation has been suggested to have a didactic role in clarifying distinctions between vowels for non-native speakers [14]. The lack of change in pitch characteristics in L2-directed speech is also in accordance with previous findings [14], suggesting that pitch adaptations have an affective role which is relevant for infant- and pet-directed speech but not foreigner-directed speech.

Even though talkers had greater communicative difficulty when interacting with a non-native talker than in the acoustic-barrier conditions, they reduced their articulation rate and increased their vowel F1 and F2 ranges significantly less in FDS than in the VOC condition, while similar minimal adaptations were obtained for mean energy, median F0 and F0 range in L2 and VOC. This lower degree of acoustic-phonetic enhancement was also reflected in lower perceived-clarity ratings for FDS. This is further evidence that listener-directed clear speaking styles are tailored to the communication barrier that the interlocutor is facing [5,7] and that the degree of acoustic enhancement in clear speech is not simply a function of the degree of communicative difficulty experienced within the interaction. Further work is needed to investigate the role of linguistic strategies in foreigner-directed speech, involving prosodic, lexical or syntactic changes.

ACKNOWLEDGMENTS

This work was supported by the Economics and Social Research Council [RES-062-23-0681]. We thank Sharna Allen for her contribution to the study.

7. REFERENCES

[1] Baker, R., Hazan, V. 2011. DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods* 43, 761-770

[2] Boersma, P., Weenink, D. 2014. Praat: doing phonetics by computer, http://www.praat.org.

[3] Bradlow, A. R., Bent, T. 2002. The clear speech effect for non-native listeners. *J. Acoust. Soc. Am.* 112, 272-284.

[4] Cooke, M., King, S., Garnier, M., Aubanel, V. 2014. The listening talker: A review of human and algorithmic context-induced modifications of speech. *Comput. Speech Lang.* 28, 543-571.

[5] Garnier, M., Henrich, N., Dubois, D., 2010. Influence of sound immersion and communicative interaction on the Lombard effect. J. *Speech Lang. Hear. Res*. 53, 588–608.

[6] Ferguson, S.H. & Kerr, E.E.. 2009. Subjective ratings of sentences in clear and conversational speech. *J. Acad. Rehabil. Audiol*. 42, 51-66.

[7] Hazan, V., Baker, R. 2011. Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *J. Acoust. Soc. Am.* 130, 2139-2152.

[8] Lam, J., Tjaden, K. & Wilding, G. 2012. Acoustics of clear speech: Effect of instruction. *J. Speech Lang. Hear. Res.* 55, 1807–1821.

[9] Lindblom, B. 1990. Explaining phonetic variation: a sketch of the H&H Theory. In Hardcastle, W., Marchal, A. (eds), *Speech Production and Speech Modelling.* Dordrecht: Kluwer, 403-439.

[10] Pinheiro J., Bates D., DebRoy S., Sarkar D, R Core Team 2014. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-117. http://CRAN.R-project.org/package=nlme.

[11] Scarborough, R., Brenier, J., Zhao, Y., Hall-Lew, L., Dmitrieva, O. 2007. An acoustic study of real and imaginary foreigner-directed speech. *Proc. 17th ICPhS*, Saarbrueken, 2165-2168.

[12]Scarborough, R., Zellou, G. 2013. Continua of clarity: “clear” speech authenticity and lexical neighborhood density effects in production and perception. *J. Acoust. Soc. Am.* 134, 3793-3807.

[13] Smiljanić, R., Bradlow, A. 2009. Speaking and hearing clearly: talker and listener factors in speaking style changes. *Language and Linguistics Compass*, 3/1, 236-264.

[14] Uther, M., Knoll, M., Burnham, D. 2007. Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant- directed speech. *Speech Comm*. 49, 2–7.

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