Vowel shortening in altered speech modes elicited via an interactive task

Joanna Sankowska[1], Maria Luisa Garcia Lecumberri[1] and Martin Cooke[1, 2]

[1] LASLAB, Depto. Filologia Inglesa, Universidad del Pais Vasco, Spain; [2] Ikerbasque
j.sankowska@hotmail.com, garcia.lecumberri@ehu.es, m.cooke@ikerbasque.org

ABSTRACT

Clear speaking styles have been demonstrated to enhance certain acoustic-phonetic features of speech. However, previous work on speech production changes has mainly used interactions with imaginary interlocutors or read speech. In this study, two types of altered speech modes were elicited through a new combination of 'gap' and 'map' tasks, which leads to unscripted dialogues between interlocutors. British speakers completed a task in cooperation with an adult native speaker or adult foreigner, in quiet and in noisy conditions. Speaker productions were analyzed to examine changes related to vowel and consonant duration in voiced and voiceless rhymes in naturally elicited speech in adverse conditions. Vowel shortening in voiceless contexts was present. The durational contrast between voiced and voiceless rhymes was reduced in altered speech styles for vowels and maintained for plosives.

Keywords: spontaneous speech, altered speech, communicative task, vowel shortening.

1. INTRODUCTION

Communication is an important factor in intelligibility-enhancing speech styles. Speech directed to the hearing-impaired or foreigners, or speech produced in noise, are governed to some extent by feedback from the interlocutor [11, 2]. Many previous studies of altered speech styles have been based on material elicited using simulated interactions where participants were asked to speak to an imagined interlocutor [3, 10] or by using carefully prepared read speech [9, 10] which does not involve communication. Different types of speech (e.g. read vs. spontaneous) can differentially influence the outcome of measured behavioural responses [11, 17]. Indeed, some phenomena cannot be properly investigated in non-communicative settings [14, 15]. Although tightly-controlled elicitation techniques offer many advantages, they lack communicative aspect of conversational speech. Speech elicited unnaturally is not ideal for investigating many of the features of altered speech styles that are present in order to improve communication. Accordingly, the present study involves interactions with real interlocutors and spontaneous speech elicited during a communicative task in situations that are adverse due to the background noise or insufficient linguistic knowledge of the interlocutor. We investigate possible interactions between temporal changes occurring in altered speech styles and duration-based differences in English rhymes which accompany consonant voicing contrasts. We also investigate the possibility that altered speech styles may enhance or degrade important cues to various features, e.g. voicing. Specifically, we focus on differences in vowel and consonant duration due to the voicing of the rhyme where they occur [5, 7, 16]. The speech styles tested were native adult-directed (ADS), foreigner-directed (FDS), and Lombard speech (LS).

2. METHOD

2.1. Speakers

Nine native speakers of Standard British English were recruited amongst the students and staff of the Department of Psychology at the University of Bristol, Bristol, UK. They were recorded in interactions with a British adult interlocutor and a foreign (Chinese) adult interlocutor with a noticeable foreign (Mandarin) accent. None of the participants had a known history of speech or hearing impairment.

2.2. Task

Pairs of participants were involved in a communicative map task designed for the purpose of this study. The original HCRC Map Task [13] uses a set of simple maps with labeled features and a route. In our study participants were given a map of an imaginary city which was much more complex than the HCRC maps. The map we designed had some gaps for missing street names and included landmarks (e.g. school, takeaway, etc.). Additionally, one of the participants had a

route from point A to point B marked on his/her map. The missing street names were the target words. Moreover, the route led through streets whose names were also target items. In order to avoid repetitiveness, three further versions of a map were created that consisted of mirror images of the original map, resulting in four different maps of identical complexity. In the first phase, speakers were instructed to talk to each other in order to find the missing street names and fill in the gaps on the maps. In the second part, the speaker with the map containing the route was asked to describe it to the listener, whose task was to follow the speaker's directions and mark the route from point A to point B on his/her map. After completing the task, participants swapped roles and completed the task again using different sets of maps. This procedure was the same for all participants. However, only native speech material was included in the analysis.

2.3. Speech material

Target items were monosyllabic English words, including both real words such as 'Court, Beat' and words with slightly altered spelling to reflect a possible street name e.g. 'Dobb, Ligg'. They were chosen to form minimal pairs that differed only in the voicing of the coda plosive and contained a long vowel /iː/, /ɔː/, /ɑː/ or a short vowel /ɪ/, /ɒ/, /æ/, followed by either a voiceless (/p/, /t/, /k/) or voiced plosive (/b/, /d/, /g/). The onset consonants agreed within each minimal pair. All target words were followed by the word 'Avenue' or 'Street' in order to facilitate manual segmentation. The maps also included filler words that were real street names e.g. 'Dove Lane, Beam Drive'. Here we present only the data collected during the second part of the task i.e. the description of the route marked on the maps. Vowels and consonants in coda position in the target words were measured for the analysis. There were 469 tokens collected in ADS, 460 in LS, 410 in the first FDS session and 359 in the second but only part of the data was used for the analysis (see 'Results' section).

2.4. Data collection

Participants were tested in a sound-attenuating booth. They worked in pairs but were separated by an opaque screen to prevent the use of visual cues in the task that might interfere with altered speech styles. All participants were recorded in 3 sessions: (i) in quiet; (ii) with speech shaped noise delivered through headphones; and (iii) in quiet with an adult foreign interlocutor, with session order counterbalanced across speakers. Recordings used head-mounted microphones (Sennheiser MZA 900 P) and an M-Audio MobilePre audio interface.

3. RESULTS

Target items were segmented using Praat [4] and vowel and coda durations measured. Durational contrasts were calculated as the difference between durations in the voiced and voiceless plosive contexts. Outliers defined on the basis of vowel duration and words without a counterpart were discarded from the analysis. The items were paired up on the basis of the number of collected minimal pairs and the contrast was calculated within each minimal pair. Native interactions resulted in each participant describing the route once and filling the gaps twice. To compensate for that in native-foreigner interactions we used 2 different maps which caused the route part being described twice. Hence there are more items collected in FDS than in ADS and LS. Additionally, we found no difference between 2 FDS parts so the data was pooled. As a result 663 tokens in total were used for the statistical analysis (173 in ADS, 321 in FDS and 169 in LS).

3.1. Vowel shortening

In this section we present segmental durations in the three speech styles analyzed separately. Table 1 shows mean vowel and plosive durations in the 3 speech styles.

	Intrinsically short vowels		
	voiceless	voiced	difference
ADS	134	151	17
FDS	134	154	20
LS	153	169	16
	Intrinsically long vowels		
	voiceless	voiced	difference
ADS	195	234	39
FDS	207	235	28
LS	224	247	23
	Plosives following intrinsically short vowels		
	voiceless	voiced	difference
ADS	132	85	47
FDS	133	83	50
LS	114	75	39
	Plosives following intrinsically long vowels		
	voiceless	voiced	difference
ADS	117	84	33
FDS	113	77	36
LS	101	76	25

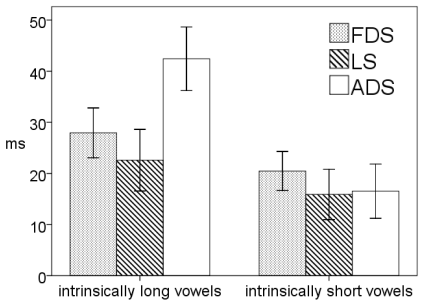
Table 1: Mean vowel and consonant durations in adult-directed, foreigner and Lombard speech (ms).

Analyses revealed similar patterns in all speech styles. Vowel shortening was statistically significant in all conditions ($p<.05$, with $p<.001$ in most cases). The duration of long and short vowels was decreased when they were followed by voiceless plosives as opposed to voiced ones. Analysis of plosives also confirmed durational differences. Voiced plosives were shorter than voiceless ones following both long and short vowels. These results suggest that vowel shortening due to the voicing properties of the following plosive was observed for both long and short vowels regardless of the speaking style.

3.2. Interaction between vowel duration and speaking style.

Here we present the results of vowel durational contrasts with two factors of length (intrinsically short vs. intrinsically long vowels) and style (ADS vs. FDS vs. LS) by mixed effects modelling with length and style as fixed factors and subject as a random factor. Figure 1 shows mean vowel durational differences in the three speaking styles. The effect of length was significant ($p<.01$) i.e. there were different tendencies for long and short vowels. For the long vowels, the durational contrast was significantly reduced in LS relative to ADS ($p<.05$). There was also a tendency for the contrast in FDS to be reduced relative to ADS but the result fell short of significance ($p=.069$). As for the short vowels, there was no difference between the conditions. These results indicate that, at least for the long vowels, the duration-based voicing contrast is reduced in Lombard speech compared to ADS. Further analysis of the ADS data revealed that the contrast was significantly smaller for short vowels than it was for long vowels ($p<.01$). No difference was found either in LS or in FDS between the long and short vowels.

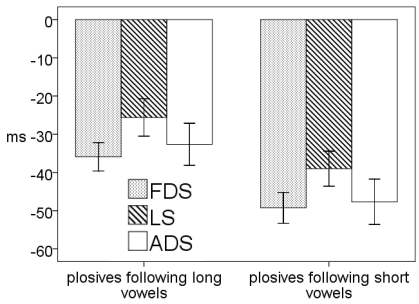
Figure 1: Mean vowel durational contrast.



3.3. Interaction between plosive duration and speaking style.

Figure 2 shows mean durational differences between voiceless and voiced consonants in the three speech styles. The effect of length was significant ($p<.001$). For LS and FDS the durational difference was reduced when the consonants followed long vowels relative to consonants following short vowels ($p<.05$). A similar trend exists for ADS ($p=.067$). For the durational differences between the styles we found similar results to those of read speech i.e. there was a tendency for the durational difference in LS to be reduced and for it to be enhanced in FDS relative to ADS. However, those results did not reach significance suggesting that in conversational speech the consonant durational contrast for FDS and LS is maintained when compared to ADS.

Figure 2: Mean plosive durational contrasts.



4. DISCUSSION

Several previous studies have examined altered speech styles using speech elicited through a set of carefully formulated instructions e.g. read speech. However, this type of speech lacks the communicative component which is important in speech styles that partly rely on the feedback from the listener. For instance, it is an important factor in eliciting Lombard speech. It is known that communicative aspects can amplify the changes from quiet to noise [6, 8]. Not only the type of adverse conditions [9] but also the type of speech elicited (read vs. spontaneous) [11, 14] can affect the outcome. Our study provided a communicative task that elicited spontaneous speech in potentially difficult interactive situations. Additionally, the study involved real rather than imaginary listeners who provided feedback to the speaker. We focused on investigating vowel shortening as a correlate of coda consonant voicing and the vowel and plosive durations related to it.

First, we confirmed that vowel shortening was present in all speaking styles. Regardless of speech mode, speakers produced shorter vowels when followed by voiceless plosives as opposed to voiced ones, and produced shorter voiced than voiceless plosives, in line with previous studies [12]. This finding confirms for natural, spontaneous speech effects which were reported in [9] for read speech.

Second, we investigated how durational contrasts were influenced by listener-directed speech styles. The comparison of speech changes revealed that significant differences between styles related only to intrinsically long vowels. These results are similar to [9], which also found that only intrinsically long vowels were affected by the different speech styles. However, our findings differ from [9] in the results for FDS where the contrast was attenuated while it was enhanced in [9]. This may be due to the fact that our study used spontaneous speech rather than read speech. This view is supported by other studies that also found a decrease in word duration in spontaneous speech as opposed to read [11, 1]. Also [17] found that durational differences between word-initial and word-final consonants were smaller in spontaneous speech as opposed to read speech.

Another possibility is that FDS itself introduces variability and is 'foreigner-dependent', i.e. different interlocutors may elicit different degrees of FDS. Additionally, we found that only in ADS there was a difference in the behaviour of the vowels i.e. the durational contrast was significantly smaller for short vowels than it was for long vowels. Analysis of the plosives showed that this contrast was attenuated for LS relative to ADS but there was no difference between FDS and ADS. This result is partly in opposition with other studies such as [9] where this contrast was enhanced for FDS relative to ADS.

5. CONCLUSIONS

Our findings suggest that the magnitude of the changes in some acoustic-phonetic phenomena (e.g. duration-related processes) varies between read and spontaneous speech. Still, the extent to which the observed changes are due to the task used or the inter-speaker variability or indeed other cues used by speakers is a matter for future study. Also, the effect of different foreigner interlocutors on FDS is worth investigating. Although our study suggests that increased intelligibility of altered speech styles is not due to the enhancement of duration-based contrasts, it is possible that other cues compensate for it and rely on different strategies employed by different speakers.

6. REFERENCES

[1] Baker, R., Hazan, V. 2009. Acoustic-phonetic characteristics of naturally-elicited clear speech in British English. *157th Meeting Acoust. Soc. Am. Portland, Oregon.*

[2] Baker, R., Hazan, V. 2010. Acoustic characteristics of clear speech produced in response to three different adverse listening conditions. *Psycholinguistic Approaches to Speech Recognition in Adverse Conditions, Bristol.*

[3] Biersack, S., Kempe, V., Knapton, L. 2005. Fine-tuning speech registers: a comparison of the prosodic features of child-directed and foreigner-directed speech. *Proc. 9th Interspeech, Lisbon*. 2401–2405.

[4] Boersma, P., Weenink, D. 2005. Praat: Doing Phonetics by Computer (Version 4.3.04) [Computer Program], Available online: <http://www.praat.org>

[5] Cruttenden, A. 2001. *Gimson´s pronunciation of English*. Oxford: A Hodder Arnold Publication.

[6] Garnier, M., Henrich, N., Dubois, D. 2010. Influence of sound immersion and communicative interaction on the Lombard effect. *J. Speech. Lang. Hear. Res*. In press.

[7] Gimson, A. C. 1989. *An introduction to the pronunciation of English*. London: Routlege.

[8] Junqua, J. C., Fincke, S., Field, K. 1999. The Lombard effect: A reflex to better communicate with others in noise. *ICASSP*. 4. 2083–2086.

[9] Sankowska, J., Cooke, M., Garcia Lecumberri, M.L. 2010. Interaction of intrinsic vowel and consonant durational correlates with Foreigner-directed speech. *Proc. 6th New Sounds, Poznan***.**

[10] Smiljanic, R., Bradlow, A. R. 2005. Production and perception of clear speech in Croatian and English. *J. Acoust. Soc. Am.* 118. 1677-1688.

[11] Spilková, H., van Dommelen, W. 2010. English of in L1 and L2 speakers' read and spontaneous speech. *Proc. Fonetik 2010, Lund, Sweden.* 91-96.

[12] Summers, W.V. 1987. Effects of stress and final-consonant voicing on vowel production: articulatory and acoustic analyses. *J. Acoust. Soc. Am.* 82. 847-863.

[13] Thompson, H. S., Anderson, A., Gurman Bard, E., Doherty-Sneddon, G., Newlands, A., Sotillo, C. 1993. The HCRC Map Task corpus: natural dialogue for speech recognition*. Proc. The workshop on Human Language Technology, Princeton, New Jersey*.

[14] Torreira, F., Adda-Decker, M. Ernestus, M. 2010. The Nijmegen corpus of casual French. *Speech Communication* 52, 201-212.

[15] Torreira, F., Ernestus, M. 2011. Vowel elision in casual French: the case of vowel /e/ in the word c'était. *Journal of Phonetics* 39, 50-58.

[16] Wells, J. C. 1981. *Accents of English.* London: Longman.

[17] White, L., Wiget, L., Rauch, O., & Mattys, S.L. 2010. Segmentation cues in spontaneous and read speech. *Proc. 5th Speech Prosody, Chicago*.