ACOUSTIC-PHONETIC CHARACTERISTICS OF CLEAR SPEECH IN BILINGUALS

Sonia Granlund, Rachel Baker & Valerie Hazan

Department of Speech, Hearing and Phonetic Sciences, University College London, UK
sonia.c.granlund@gmail.com; rachelbaker81@gmail.com; v.hazan@ucl.ac.uk

ABSTRACT

This study examined the language-dependency of clear speech modifications by comparing the clear speech strategies of late bilinguals in both their L1 (Finnish) and L2 (English). Results generally supported the hypothesis of language-independent enhancement of global clear speech modifications, but language-dependent segmental enhancement. The global clear speech strategies produced by Finnish-English bilinguals in their L2 (English) were similar in the extent of the modifications to those of native English speakers, indicating a surprising flexibility of the non-native speech production system.

Keywords: clear speech, L1-L2 interaction, L2 speech production

1. INTRODUCTION

Clear speech research has revealed a wide range of acoustic-phonetic modifications made by speakers when attempting to clarify their speech to the listener. However, most studies have assessed clear speech production in native English speakers, and it is therefore uncertain whether the type and extent of clear speech modifications previously found also occur in other languages and populations.

It has been proposed that so-called global modifications – a wider F0 range, higher mean F0, slower speech rate and increased speech intensity – enhance the overall salience of the speech to make it generally more perceptible for listeners. These clear speech adjustments would therefore arise independently of the language being spoken [2]. Segmental modifications – such as increased VOT differences between voiced and voiceless stops and increased spectral distances between front and back vowels – are argued to reflect the greater approximation of phonetic targets, with the aim of making the phonological categories more distinct for listeners [7]. Such clear speech adjustments would be language-dependent, reflecting contrastive categories of the language. A few studies have assessed this hypothesis by comparing clear speech strategies across groups of speakers cross-linguistically [8, 10]. However, as enhancement strategies vary greatly between speakers [3], group differences may reflect individual variability rather than language effects.

Another issue is whether non-native speakers are able to produce clear speech in their second language – it is predicted that L2 speakers can produce language-independent global clear speech modifications, but that they would have difficulty at the segmental level as they may not have acquired the correct L2 phonetic targets, especially for categories similar to those in their L1 [5]. Speakers may also be unable to modify their speech when speaking clearly due to inflexible L2 speech production mechanisms. The few studies assessing non-native clear speech production e.g. [9] have included limited global acoustic analyses.

The aim of our study was to draw together a more coherent assessment of language-dependent clear speech strategies and non-native clear speech production. The clear speech produced by late Finnish-English bilinguals was assessed in both their L1 and L2, and their L2 clear speech strategies were compared to those of native English speakers. Finnish is typologically different to the mostly Indo-European languages that clear speech research has examined so far. A wide range of acoustic-phonetic measures was included. Importantly, as clear speech is a listener-oriented speaking style [7], a more naturalistic, communicative task [6] was used to obtain global measures of pitch median and range, speaking rate and mean long-term energy. A sentence-reading task was used to acquire fine-grained segmental measures of vowel cues and VOT in more controlled environments.

2. METHODS

2.1. Participants

Twelve female, extremely proficient Finnish-English late bilingual participants (20-35 years

old) took part in the study. All had learned English as a second language at school for over 9 years. All but one lived in London at the time of recording, and, on average, had resided in an English-speaking country for 5.8 years. They reported no speech or hearing impairments.

2.2. Materials

Measurements of global measures in casual and clear speech were made on spontaneous speech recordings. The spontaneous speech was elicited using the diapix task [4], a collaborative 'spot the difference' task in which two participants have to find differences in the pictures through dialog without view of each other [1].

For the measurement of segmental features, it is necessary to control the environment in which the segments are produced, so words containing the segments under investigation were embedded in read sentences. Two segmental contrasts were evaluated for each language: voiced/voiceless bilabial plosives and a vowel contrast. As Finnish, unlike English, does not distinguish between voiced and voiceless stops, Flege's [5] speech learning model (SLM) would predict that the similar short-lag Finnish /p/ and English /b/ may be equated, leading to similar clear speech enhancement strategies across languages for these phonemes. English and Finnish sentences containing keywords differing in initial consonant were created: 4 English /p-b/ minimal pair keywords were used, and 4 Finnish keywords with initial voiceless /p/ were matched with the English keywords. Each was produced in 2 sentences per condition in English, and 4 sentences in Finnish, giving 16 VOT sentences per language. The primary cue for the vowel tense-lax distinction in English is spectral, but the temporal distinction is more important in Finnish. To investigate whether the temporal and spectral aspects of the high front vowel pair /i/-/ɪ/ (tense-lax) in English and the /i̟ː/-/i/ (long-short) distinction in Finnish are enhanced differently in clear speech, 8 further minimal pair keywords were constructed (4 per language). Cross-linguistically the vowels were matched as closely as possible for segmental context and keyword position. Each keyword was placed in two different sentences.

2.3. Procedure

For the diapix task, the recordings were done in pairs. The participants wore headsets with microphones and were seated in separate acoustic booths. They took part in two recording sessions. In each session, recordings were done in one language only. Two picture tasks were completed in the 'no-barrier' condition (NB), in which the participants could hear each other normally, thus producing casual speech. A further 4 pictures were done in the vocoder condition (VOC), in which the speech of one of the two participants was distorted by a three-channel vocoder in real time to produce a communication barrier [1]. To complete the problem-solving task successfully, the 'unimpaired' talker has to produce clear speech to be understood by the talker hearing vocoded speech. Only the recordings of the participant as the 'unimpaired' talker, i.e. the partner producing clear speech, were analysed. Overall, speech was analysed from 2 pictures in the 'casual' NB condition and 2 pictures in the 'clear' VOC condition for each participant in each language.

In the sentence reading task, participants were asked to say the sentences 'as if talking to a friend' to produce a casual speaking style, and 'as if talking to someone with a hearing impairment' to produce a clear speaking style [8, 9].

2.4. Processing

The audio files of the diapix task in both languages were transcribed and aligned at word level to the waveform. The same processing procedures as in [6] were used for obtaining F0 median and range, mean energy and word duration measures.

The VOT of each keyword in the reading task was segmented manually, and the mean VOT duration for each participant in each condition was calculated. To measure the vowels, the interval was segmented manually, and the duration of the vowel interval, as well as its midpoint F1 and F2, were calculated using a script. The means for each measure for each participant were used in the analysis.

3. RESULTS

3.1. Do late bilinguals use different clear speech strategies in Finnish (L1) and English (L2)?

Repeated measures ANOVAs were carried out on each of the measures of median F0, F0 range, mean energy 1-3kHz and mean word duration with language (L1 Finnish, L2 English) and condition (NB, VOC) as within-subject factors. Table 1 displays the means for each measure, and table 2

summarizes the statistical analyses. Main effects of condition show that F0 median, mean energy and word duration increased from the casual (NB) to the clear speech (VOC) conditions but there was no effect of language for F0 median and mean energy. The only interaction of language and condition, indicating a difference in clear speech strategies across languages, was for mean word duration: in clear speech, mean word duration increased more in Finnish than in English.

Table 1: The means of F0 median, F0 range, mean energy (ME) and mean word duration (MWD) in the 'casual' NB and 'clear' VOC conditions for the late bilinguals (L1 Fin, L2 Eng) and native English speakers (N Eng).

diapix_NB	L1 Fin	L2 Eng	N Eng
f0 median (Hz)	188.4	188.7	199.2
f0 range (Hz)	48.2	37.8	37.0
ME (dB)	26.1	26.5	26.7
MWD (ms)	315.6	307.9	254.8
diapix_VOC	L1 Fin	L2 Eng	N Eng
f0 median (Hz)	196.4	199.8	210.7
f0 range (Hz)	43.4	40.5	42.7
ME (dB)	28.7	29.0	28.1
MWD (ms)	395.9	369.3	338.3
% change	L1 Fin	L2 Eng	N Eng
f0 median (Hz)	4.5	6.4	5.8
f0 range (Hz)	1.4	10.8	16.0
ME (dB)	9.0	9.5	5.2
MWD (ms)	26.0	20.6	33.4

Next, segmental enhancements were considered. An ANOVA was run on the VOT durations of English /b/ and Finnish /p/ with within-subjects factors of segment (b, p) and condition (casual, clear). There was a significant interaction between segment and condition [$F(1,10)=5.9$; $p<0.05$]: English /b/ and Finnish /p/ did not differ in VOT in the casual speech condition. In clear speech, VOT decreased relative to the casual speech condition, but the VOT for Finnish /p/ decreased less (17 ms to 14 ms) than that of English /b/ (16 ms to 11 ms).The results therefore indicate a difference in enhancement strategies across languages. To explore whether the speakers enhance the VOT of the English long-lag /p/, a paired t-test was performed on the casual and clear tokens. VOT was greater in the clear (79 ms) than in the casual condition (63 ms) [$t=-2.39$; $p<0.05$; $df=10$]: bilinguals were able to enhance the 'new' VOT category in the opposite direction to the short-lag stops.

Table 2: The results of repeated measures ANOVAs on the global measures as produced by the bilingual speakers in Finnish and English.

		F	df	p	
median F0	language	0.62	(1,9)	0.45	
	condition	6.33	(1,9)	0.033	*
	interaction	0.52	(1,9)	0.488	
F0 range	language	5.02	(1,9)	0.052	
	condition	0.03	(1,9)	0.857	
	interaction	2.26	(1,9)	0.167	
mean energy 1-3kHz	language	0.14	(1,5)	0.723	
	condition	41.07	(1,5)	0.001	**
	interaction	0.86	(1,5)	0.395	
mean word duration	language	5.48	(1,11)	0.039	*
	condition	23.93	(1,11)	<0.001	**
	interaction	6.82	(1,11)	0.024	*

As a measure of the amount of spectral enhancement for each of the vowels, the F1/F2 Euclidean distance was calculated for each vowel. An ANOVA was run on the measure with within-subjects factors of language (L1 Finnish, L2 English), segment (short, long) and condition (casual, clear). The vowels had a greater F1/F2 spectral distance in the clear condition than in the casual condition [$F(1,10)=23.89$; $p<0.001$]. However, neither the interaction between language and condition nor the three-way interaction were significant, and therefore the strategy for spectral enhancement was not different in the English vowels as compared to the Finnish vowels.

To explore whether speakers' specific spectral enhancement strategies nevertheless differ in the two languages, an ANOVA was run separately for the F1 and F2 measures with within-subjects factors of language (L1 Finnish, L2 English), segment (short, long) and condition (casual, clear). For F1, a significant interaction of language and condition was found [$F(1,10)=8.6$; $p<0.05$], suggesting that the speakers use different strategies for enhancing vowel F1 in the two languages. T-tests indicate that the F1 of English vowels is not decreased from the casual condition to the clear condition (423Hz to 413Hz), but the F1 of Finnish vowels is lowered (391Hz to 348Hz). Vowels were found to have a higher F2 in clear speech than in casual speech (2735Hz vs. 2596Hz) [$F(1,10)=25.9$; $p<0.001$], but there were no interactions, suggesting similar strategies across languages.

An ANOVA was run on the duration of each vowel, with within-subjects factors of language (L1 Finnish, L2 English), segment (short, long)

and condition (casual, clear). The vowels in the clear condition were significantly longer than in the casual condition [$F(1,10)=65.41$; $p<0.001$], but there was no interaction of language and condition or a three-way interaction: the Finnish vowels were not lengthened more than the English vowels, contrary to predictions.

3.2 Do the global characteristics of the clear speech of late bilinguals in English differ from that of native speakers?

The global measures for the L2 English of the Finnish-English late bilinguals from the diapix task were compared to similar materials for 12 native English speakers randomly chosen from the LUCID database [1]. As a more reliable comparison across speakers, the measure of percent change from the NB to the VOC condition for median F0, F0 range, mean word duration and mean energy was calculated (see table 1). An independent samples t-test then examined the effect of native language (Finnish, English) on each measure. There were no significant differences in the extent of the clear speech modifications made for any of the measures, suggesting that the late bilingual speakers were able to modify their speech on these dimensions similarly to native speakers.

4. DISCUSSION

Our data generally supports the hypothesis of language-independent enhancement of global measures but language-dependent enhancement of segmental cues. For global modifications, only speech rate, measured in word duration, was found to be modified differently in Finnish and English – this may have arisen from structural differences between the words of the two languages.

In clear speech, the VOT for the English short-lag /b/ decreased more than for the Finnish short-lag /p/, and speakers also enhanced the 'new' aspirated English /p/ category by modifying VOT in the opposite direction. This indicates that segmental modifications are guided by contrast enhancement – speakers are attempting to increase the acoustic distance between the 'voiced' and 'voiceless' categories in English. The language-specificity was not as evident in the vowels – the extent of spectral and temporal enhancement was similar in Finnish and English, despite different primary vowel cues. However, the F1 of English vowels was not lowered in clear speech, suggesting that the spectral distances between the two English high vowels may be maintained.

The late bilinguals were able to produce global clear speech modifications in their L2 to a similar extent to native speakers. This also implies that non-native speakers are able to adapt their speech according to the listener's needs as native speakers do, indicating surprisingly flexible L2 speech production mechanisms.

5. ACKNOWLEDGMENTS

This work was partly funded by the UK Economics and Social Research Council (RES-062-23-0681). Thanks to Martti Vainio and colleagues for their help with the automatic alignment in Finnish.

6. REFERENCES

[1] Baker, R, Hazan, V. In press. DiapixUK: a task for the elicitation of spontaneous speech dialogs. *Behavior Research Methods*.

[2] Bradlow, A.R., Bent, T. 2002. The clear speech effect for non-native listeners. *J. Acoust. Soc. Am.* 112(1), 272-284.

[3] 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 and Speech 53(4), 510-540.

[4] Ferguson, S.H., Kewley-Port, D. 2007. Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *JSLHR* 50, 1241-1255.

[5] Flege, J.E. 1987. The production of 'new' and 'similar' phones in a foreign language: Evidence for the effect of Equivalence Classification. *J. Phon* 15, 47-65.

[6] Hazan, V., Baker, R. Conditionally accepted. Acoustic-phonetic characteristics of clear speech produced with and without communicative intent. *J. Acoust. Soc. Am*.

[7] Lindblom, B. 1990. Explaining phonetic variation: a sketch of the H&H theory. In Hardcastle, W.J., Marchal, A. (eds.), *Speech Production and Speech Modeling*. Amsterdam, The Netherlands: Kluwer Academic, 403-439.

[8] Smiljanic, R., Bradlow, A.R. 2005. Production and perception of clear speech in Croatian and English. *J. Acoust. Soc. Am.* 118(3), 1677-1688.

[9] Smiljanic, R., Bradlow, A.R. 2009. Native and non-native clear speech production. *J. Acoust. Soc. Am*. 125(4), 2753.

[10] Wassink, A., Wright, R., Franklin, A. 2007. Intraspeaker variability in vowel production: An investigation of motherese, hyperspeech, and Lombard speech in Jamaican speakers. *J. of Phon*, 35, 363-379.