Title: An acoustic-phonetic comparison of the clear speaking styles of Finnish-English late bilinguals [a)]

Author names and affiliations:

Sonia GRANLUND [b], Valerie HAZAN [c] and Rachel BAKER [d]

[b] UCL Speech Hearing and Phonetic Sciences, Chandler House, 2 Wakefield Street, London WC1N 1PF, UK. Email: sonia.c.granlund@gmail.com

[c] UCL Speech Hearing and Phonetic Sciences, Chandler House, 2 Wakefield Street, London WC1N 1PF, UK. Email: v.hazan@ucl.ac.uk

[d] UCL Speech Hearing and Phonetic Sciences, Chandler House, 2 Wakefield Street, London WC1N 1PF, UK. Email: rachelbaker81@gmail.com

Authors version of a paper accepted for publication at the Journal of Phonetics (20 February 2012). doi:10.1016/j.wocn.2012.02.006.

---

a) Part of this work was presented in "Acoustic-phonetic characteristics of clear speech in bilinguals" at the International Congress of Phonetic Sciences (ICPhS), Hong Kong, 17-21 August 2011

Abstract

Research on clear speech, an intelligibility-enhancing speaking style, has proposed that global clear speech modifications which make speech more perceptible in adverse conditions are language-independent, while the more fine-grained segmental clear speech modifications, which enhance the salience of phonological contrasts, are language-specific [Bradlow, A.R. & Bent, T., 2002. The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America*, *112*, 272–284]. This study assessed the claim by contrasting the clear speech strategies used by twelve Finnish-English late bilinguals in their two languages, using spontaneous speech and sentence reading tasks. Their global clear speech modifications were also compared to those of native English speakers. Global measures included mean energy between 1-3k Hz, f0 median and range, and speech rate, while segmental measures included VOT for initial stop consonants and spectral and temporal characteristics for two vowel contrasts. Findings generally support the hypothesis that global enhancements are language-independent: most of the global clear speech modifications were consistent across languages. However, segmental enhancements were not consistently language-dependent: the late bilinguals enhanced stop voicing contrasts according to the language being spoken, but vowels were modified similarly in the clear speaking style of both languages. The global clear speech strategies of late bilinguals were found to approximate those of native English speakers.

Keywords: clear speech, bilingual speech production.

1. Introduction

Late bilinguals often achieve high levels of speech fluency in their second language, even if they retain a foreign accent (Munro & Derwing, 1995; 1999). A vital aspect of speech fluency is the ability to adapt one's speech to different communicative situations so as to be more intelligible to the person with whom we are interacting (Lindblom, 1990). The present study involves the elicitation of clear speech strategies in Finnish-English late bilinguals in both their languages. It assesses the extent to which these clear speech strategies are language-specific and the extent to which they approximate the strategies used by monolingual speakers. These questions are important as they can inform us about how we use the control we have over our speech production to maximise the effectiveness of our speech communication. They also inform us about the factors that guide clear speech production and the associated increase in intelligibility (Smiljanic & Bradlow, 2005; 2009a).

Clear speech is a speaking style which speakers adopt when their interlocutor has difficulty understanding them due to background noise, hearing impairment, lack of experience in the language or insufficient linguistic context (for a review, see Smiljanic & Bradlow, 2009a). In such situations, Lindblom's (1990) Hyper-Hypo (H&H) theory predicts that speakers increase articulatory effort (hyperarticulate) to ensure successful communication, whereas they apply as little effort as possible in speech when there are no communication difficulties and where the content is highly predictable (hypoarticulation). Studies have shown that clear speech enhances intelligibility for both normal-hearing and hearing-impaired listeners (e.g., Picheny et al., 1985; Liu et al., 2004), and research has established a wide range of acoustic-phonetic modifications typically made in English clear speech. So-called global changes include a decrease in speech rate, an increase in the frequency and duration of pauses, in pitch range and mean fundamental frequency, and an amplification of the 1-3k Hz frequency region (e.g., Picheny et al., 1986; Krause & Braida, 2004; Liu et al., 2004; Smiljanic & Bradlow, 2005). Clear speaking styles also involve changes at the segmental level. Indeed, English clear speech modifications have been shown to involve more frequent releasing of word-final consonants (Picheny et al., 1986; Bradlow et al., 2003; Krause & Braida, 2004), an increase in duration contrasts between the voice-onset-time (VOT) of voiced and voiceless stops (e.g., Smiljanic & Bradlow, 2008) and between "tense" and "lax" vowels (Uchanski, 1988), an increased spectral distance between front and back vowels (Ferguson & Kewley-Port, 2002), and a more expanded vowel space (e.g., Moon & Lindblom, 1994; Bradlow, 2002; Krause & Braida, 2004).

It has been argued that global and segmental adjustments in clear speech serve different purposes. Global adjustments increase the overall salience of the signal so that the speech is generally more perceptible and audible (Bradlow & Bent, 2002). On the other hand, clear speech modifications at the segmental level reflect the greater approximation of phonetic targets, with the aim of making the phonological categories of the language more distinct (Lindblom, 1990; Johnson et al., 1993). We can hypothesise that if global adjustments aim to enhance the overall audibility of the signal, the same global clear speech modifications should occur independent of the language being spoken. On the other hand, given that languages vary in their contrastive categories, segmental enhancement strategies are expected to be language-dependent.

To our knowledge, only two studies have explored global enhancements cross-linguistically, with different groups of native speakers in each language; speakers of English and Croatian decrease their speech rate and increase the f0 range equally in clear speech (Smiljanic & Bradlow, 2005), and f0 and

intensity modifications for Jamaican Creole and Jamaican English hyperspeech were found to be similar (Wassink et al., 2007). Global enhancements found in English, nevertheless, are not necessarily universal to all languages, or may not necessarily enhance overall salience. Intonation patterns, for example, may not be as important for conveying meaning in other languages as in English, and therefore the increased f0 mean and range may not be a universal characteristic of clear speech. Indeed, Cho et al. (2011) found that in Korean, a language which does not make use of lexical stress or pitch accent, f0 range, peak and minimum did not change from casual to clear speech. Therefore, some global enhancement strategies may instead be language-dependent and reflect the prosodic characteristics of a language (Smiljanic & Bradlow, 2009a). Further investigations are needed with languages that are dissimilar to English in terms of their prosodic and temporal patterns.

A few studies have investigated whether segmental enhancements are language-specific by comparing clear speech strategies in two different languages, again using native speakers in each language. English and Croatian use different VOT cues to voiced-voiceless stop contrasts, and this was reflected in speakers' clear speech strategies; in English, the difference in VOT between voiced and voiceless stops was enhanced by lengthening the aspiration of voiceless stops, while in Croatian the prevoicing of the voiced category was increased (Smiljanic & Bradlow, 2008). In Korean, there was evidence that older speakers, who use VOT as a primary cue to distinguish between stop voicing contrasts, enhanced that dimension in clear speech while younger speakers, with mainly an f0-based distinction, primarily increased the f0 differences (Kang & Guion, 2008). These findings strongly suggest that clear speech involves language-specific enhancements of phonetic targets. Other cross-linguistic findings, however, have shown less support for this hypothesis. Vowel spaces in clear speech have been found to expand equally despite differences in the sizes of vowel inventories of different languages (Smiljanic & Bradlow, 2005; Bradlow, 2002). It could be expected that a more crowded vowel space would require more expansion due to greater vowel confusability. Similarly, no differences in enhancement strategies were found between speakers of languages which have primary vowel cues that are either temporal or spectral, although the hypothesis predicts that vowels are enhanced according to the primary cue in the language (Wassink et al., 2007; Smiljanic & Bradlow, 2005).

Thus there is conflicting evidence as to whether both global and segmental features are language-universal or language-specific. This may at least partly be due to the fact that, in cross-language studies, different speakers were recorded in each language. This is problematic as the degree and types of enhancements shown in clear speech may vary widely between individuals (e.g., Ferguson & Kewley-Port, 2007). Between-group differences in enhancement strategies may therefore reflect individual speaker strategies rather than language-based effects; this is a serious concern in studies involving small numbers of speakers, e.g., five per language in Smiljanic & Bradlow (2005; 2008) and Wassink et al. (2007). One way of overcoming this difficulty is to evaluate clear speech strategies in a group of bilingual speakers who are highly proficient in both their languages. This is the approach taken in our study.

To our knowledge, only one previous study has examined the clear speech strategies of bilingual speakers (Bradlow, 2002), and it included limited acoustic analyses. However, some studies have investigated whether second-language learners, typically of intermediate proficiency in their non-native language, are able to clarify their speech in the non-native language. L2 speakers do not

produce stylistic variation such as vowel reduction or intonational changes according to situation formality, as native speakers do (Ulbrich, 2008; Gut, 2006). An increased cognitive load and inexperience in the cues less susceptible to distortion in the L2 may also affect the speakers' ability to modify their speech effectively (Bradlow & Alexander, 2007). The clear speech produced by late learners of English has indeed been found not to benefit native English speakers to the same extent as clear speech produced by native speakers (Li, 2009; Rogers et al., 2010). It is possible, therefore, that less proficient non-native speakers cannot employ the phonological contrast enhancements in their second language because this aspect of clear speech is language-dependent (Bradlow & Bent, 2002). It has been proposed that any benefit that non-native speakers gain from native clear speech and any intelligibility benefit their clear speech produces may be due to the language-independent global enhancement of the signal which increases the overall salience of the speech (Bradlow & Bent, 2002). However, as suggested above, it may be the case that global enhancements are also language-specific to a certain extent.

There is some evidence that increased proficiency in the second language leads to a greater ability to clarify speech in difficult communicative situations. Indeed, highly proficient non-native speakers both produce a clear speech intelligibility gain and benefit from clear speech to a similar extent to native speakers (Smiljanic & Bradlow, 2011); their increased L2 experience may lead to more accurate language-dependent clear speech strategies being used (Smiljanic & Bradlow, 2009a). An important question is whether the clear speech strategies of highly proficient non-native speakers approximate those of native speakers. On the segmental level, the clear speech production of proficient Croatian speakers of English was found to differ from that of native speakers for VOT, vowel duration before voiced and voiceless stops, and in their vowel space expansion (Smiljanic & Bradlow, 2009b). However, the clear speech productions of non-native speakers were similar to those of native speakers in the temporal distinctions between tense and lax vowels (Smiljanic & Bradlow, 2009b), and for vowel space expansion in early Spanish-English bilinguals (Bradlow, 2002). Few studies have assessed global adaptations, and limited acoustic analyses have been used; both f0 increase and speech rate decrease in clear speech were found to be similar in Cantonese late learners of English and native English speakers (Li, 2009), but proficient Croatian speakers of English were found to decrease their speech rate less than native English speakers (Smiljanic & Bradlow, 2009b).

The design of our study of clear speech strategies is novel in a number of respects. First, as our participants are highly-proficient Finnish-English late bilingual speakers tested in both their languages, the study permits within- rather than between-speaker comparisons of clear speech acoustic changes. This prevents language effects from being confounded with individual speaker effects. Second, global clear speech strategies are analysed from speech that is elicited using a more ecologically-valid task which involves spontaneous speech interactions between two speakers. The majority of previous studies have elicited clear speech using read speech, in which speakers were instructed to speak clearly. The use of a naturalistic task involving communicative intent is important when examining non-native speakers; both the accuracy of speakers' productions in their second language, and the amount of transfer from their L1 to their L2 may depend on the type of task in which the speech is elicited (Hansen, 2006; Leather & James, 1996). Finally, our study investigates speech in a language (Finnish) which is typologically quite different from the languages where clear speech has been previously studied (English, Spanish, Croatian, Korean). Finnish has a smaller vowel inventory of 8 monophthongs instead of over 10 in English (Iivonen, 1998). As a quantity language, it

has short-long vowel contrasts which differ almost exclusively in duration, whereas English has tense-lax contrasts which differ primarily in vowel spectrum and, less importantly, in duration. Finnish only implements short-lag "voiceless" stops, and therefore does not distinguish between voiced and voiceless stops as English does[1]. Prosodically, Finnish may make less grammatical use of intonation than English (Sajavaara & Dufva, 2001). It also differs from English in exhibiting primary stress on the first syllable of all words, and by being neither syllable- nor stress-timed (Iivonen, 1998).

We predict that the late-bilingual speakers will make similar modifications to the global acoustic-phonetic characteristics assessed in this study, speech rate, mean energy, f0 range and median f0, in clear speech in both their languages, while the segmental modifications to vowel cues and VOT will differ according to the language being spoken. Moreover, if the global enhancements are language-independent, we would expect the highly proficient bilinguals to produce similar global enhancements in their L2 clear speech as do native speakers.

2. Method

2.1. Participants

Twelve female native speakers of Finnish (mean age 29.1 years) took part in the study[2]. All were either university students or had completed at least an undergraduate degree. One participant was a Finnish-Swedish early bilingual, but her reported use of Swedish both currently and in her childhood was infrequent, and she had grown up in a predominantly Finnish environment. However, due to her Finnish-Swedish bilingualism, her segmental clear speech modifications were excluded from the analysis. All other participants were brought up as monolinguals. All participants had learned English as a second language at school for over 9 years. The participants came from several different regions in Finland, but only one participant spoke Finnish with a noticeable regional accent.

All participants were extremely proficient speakers of English, which was reflected in their self-reported English speaking and listening skills which all rated as either "very good" or "native-like". All but one participant lived in London at the time of recording, and their mean age of moving to an English-speaking country was 21.8 years. The participants differed greatly in the amount of time they had resided in an English-speaking country (0-15 years) but, on average, they had lived there for 5.8 years. Accordingly, they were very frequent users of English: during a typical week, 10 participants reported using English more than Finnish, one used Finnish and English equally and one spoke Finnish more than English. Although the speakers reported a wide range of influences on their accent, on average, Southern British English had been the most influential. More detailed information about the participants can be found in Appendix A.

A hearing screening test at octave frequencies between 250 Hz and 8000 Hz was conducted to test for normal hearing thresholds. Two participants had slightly elevated hearing thresholds at 8000 Hz in their left ear (25 and 35 dB HL), but all other thresholds were within normal range (20 dB HL or better). Three participants had had some form of elocution lessons, and only one participant

[1] Ringen & Suomi (2009) note that modern Finnish uses a few loan words in which word-initial voiced stops may occur, but many speakers do not implement voicing in those words, with mostly short-lag stops being used. In addition, in some contexts, /t/ may be voiced word-medially.

[2] A further 2 participants were recorded but they were excluded due to speech impairment.

reported experience of regularly communicating with a person with a hearing-impairment. The participants were not aware of the purpose of the recordings and were paid for their participation.

All participants were recorded in two different types of tasks: a spontaneous speech task for eliciting global speech measures and a sentence reading task for eliciting segmental measures from their speech.

2.2. Global measures

2.2.1. Diapix materials

The global measures were carried out on spontaneous speech produced using a collaborative problem-solving task called diapix. Originally developed by Van Engen et al. (2010), the diapix task is designed to elicit natural spontaneous speech dialogues in a laboratory setting. As the type or extent of modifications speakers make to their speech can indeed depend on whether an interlocutor is real or imagined (Charles-Luce, 1997; Scarborough et al., 2007), the elicitation of more naturalistic clear speech is an advantage, especially when examining global measures such as pitch range and speech rate. Most studies of clear speech have involved participants reading a set of sentences clearly, and then again "as if speaking to someone with a hearing impairment" or similar instruction. In the diapix task, clear speech is elicited naturally by placing a communication barrier (vocoded speech) on one speaker while they are carrying out the problem-solving task, thus eliciting the use of clear speech strategies in the other speaker.

The spontaneous speech elicited in this task was used for the measurement of the global measures in a "casual" baseline condition and in a "clear" speaking style. The previous clear speech literature refers to the baseline condition as "conversational". However, because the term implies that speakers are not participating in conversation in the clear speech condition, as is the case here with the diapix task, in the current study the term "casual" is used instead of "conversational".

In the diapix task, two participants are each given a different version of a picture scene; they are required to find the differences between the two pictures by talking to each other without view of each other. The diapixUK materials (Baker & Hazan, 2011) consist of 12 cartoon-like picture pairs, shown to be of similar difficulty, each containing twelve differences[3]. There are four picture-pairs for each of three types of scenes: beach, farm and street scenes. As the pictures contain some writing, the text in six picture pairs (two per scene type) was translated into Finnish for use in the Finnish diapix session. The remaining six picture-pairs were used in the English session. Table 1 shows an example of the type of language that is elicited in the diapix task.

[3] For example pictures, see Baker & Hazan (2011).

Table 1. A transcription of a 20-second excerpt of a conversation between two participants engaged in the diapix task. The #-symbol represents a pause or silence.

speaker A	**speaker B**
and then on the door #	
it has a sign on it and it says push #	
	mine doesn't have a sign on the door #
okay so that's number four #	
	yeah # and it's got two #
	bins # green bins that are full #
yeah #	like the bins are open because they are so full #
mine mine only has one that's full	
the other one's the the lid's closed #	
so that's number five	okay so that's a fifth yeah

2.2.2. Procedure

The diapix tasks were recorded using Adobe Audition using a sample rate of 44,100 Hz (16 bit) with EMU 0404 USB audio set-up. The participants wore Beyerdynamic DT297PV headsets with condenser cardioid microphones and they were sitting in separate acoustic booths.

The participants took part in two recording sessions. In each session, recordings were done in one language only. The participants took part in the diapix task in pairs[4]. For each recording session, the pair completed 6 diapix tasks. Two of these picture tasks were completed in the "no barrier" baseline condition (diapix_NB), in which the participants could hear each other normally. A further 4 pictures were done in the vocoder condition (diapix_VOC), aimed at eliciting a clear speaking style from one of the participants. For this condition, one of the participants' speech (the "unimpaired" talker, speaker A) was distorted by a live three-channel vocoder (Rosen et al., 1999), which divides the speech signal into three spectral bands only, and is noise-excited. This results in a significant degradation of spectral information and complete loss of the fundamental frequency, which severely lowers the intelligibility of the speech heard by speaker B, the "impaired" talker. In this condition, in order to complete the problem-solving task successfully, the "unimpaired" speaker A has to produce clear speech to be understood by the "impaired" speaker B. Vocoding has previously been used successfully to elicit a clear speaking style in participants in diapix tasks (Hazan & Baker, 2011). Vocoding was chosen for the current experiment as it simulates a cochlear implant for speaker B, and therefore the speech produced by speaker A can be compared to the speech produced in sentence reading tasks in studies which have asked participants to "read the sentences as if to someone with a hearing impairment". Hazan & Baker (2011) found that native English speakers modify their clear speech patterns differently according to the communication barrier; in the diapix_VOC condition, where f0 and intensity enhancements are unlikely to aid intelligibility, speakers did not change their f0 median and range, and mean energy and vowel F1 increased less than in a condition with a background noise barrier. Therefore in the current study we would also not expect speakers to make large changes to f0 median and range in the clear speech which is produced to counter the effects of the vocoder.

[4] Half of the participants volunteered with a friend, and the other six participants were grouped into "strangers" pairs, and thus did not know each other beforehand.

Each participant completed two pictures as the "unimpaired" speaker A with their partner as the "impaired" talker speaker B, after which they switched places, so that both participants experienced being the "unimpaired" speaker A, from whom a clear speaking style was elicited. As a significant learning effect is found for vocoded speech (e.g., Davis et al., 2005), both participants completed a 10-minute vocoder training session on a computer prior to the recordings. Previous research (Bent et al., 2009) has found that significant learning for vocoded speech occurs within the first 10 minutes; here a smaller number of channels in the vocoder was used so the learning effect may not be complete but should have diminished prior to the commencement of the recordings[5].

For the diapix task, the participants were asked to start their exploration of the differences across the pictures from the top left corner of the picture and to continue clockwise around the picture. The recording was stopped either once the participants had found all the 12 differences or once 15 minutes had elapsed.

The speech of both participants was of interest in the "no barrier" condition, and therefore both talkers were told to "try and contribute equally" to finding the differences in the pictures. In the diapix_VOC condition, only the "unimpaired" talker's speech was analysed, and therefore that speaker was asked to contribute to the discussion to a greater extent.

The diapix task was exactly the same in each of the two sessions, with the exception of the language used in the recordings; half of the pairs started with the English session, and half started with the Finnish session. The pictures for each language were semi-randomised so that each pair started with a different picture, but none of the pictures in the same condition were of the same scene. Additionally, participants completed pictures of the same scene in the same order across languages. Each pair therefore completed all 12 diapixUK pictures, and none were presented more than once. In total, approximately 12.2 hours of dual-channel diapix recordings were made, with each diapix picture taking, on average, 10 minutes to complete. After "filler" words, silences, pauses, laughter, and other non-speech sounds had been excluded, approximately 34 minutes of speech per participant, and 6.9 hours of speech overall, was analysed in the diapix task.

2.2.3. Processing

During recording, the speech produced by each participant was saved on a separate audio channel. Each channel was transcribed using Wavescroller (Northwestern University Linguistics Department software) for the audio files of both languages. The criteria used for the orthographic transcription were the same as those in Hazan & Baker (2011). For the Finnish files, the transcription was done following the same general guidelines as the English transcription. For the English files, the waveforms were aligned with the transcriptions at the word level using NUAligner software (Northwestern University Linguistics Department software), which created a Praat TextGrid (Boersma & Weenink, 2001) for each file. The Finnish word-level alignment was done using an HMM-based labeler by the Automatic Speech Recognition group at Aalto University. The alignment created text files which were converted into Praat TextGrids using a Praat script. The audio files of half of the participants were also normalised for amplitude using Adobe Audition, to a mean of 15

[5] Additionally, Hazan & Baker (2011) demonstrated that there were no differences in transaction time between the three VOC pictures which they elicited per participant. This suggests that there was no decrease in difficulty during the VOC task.

dB (with soft limiting). The audio files of the other six participants could not be normalised due to frequent high-amplitude regions of laughter occurring in the files. Several acoustic-phonetic measures were made, using the same protocol as in Hazan & Baker (2011).

2.2.3.1. Transaction time measures

As a measure of task difficulty (Van Engen et al., 2010), the time it took each pair to find the first eight differences in each condition was noted. The eighth difference was found suitable as all participants found at least eight of the twelve differences.

2.2.3.2. Fundamental frequency measures

A Praat script was run on all the audio files to obtain the fundamental frequency measures. Using a time step of 150 values per second, the script calculated the mean fundamental frequency and the interquartile range (in Hertz) of each file. For each talker, the values acquired from the recordings of the two diapix tasks in the same condition were averaged to produce a single measure of fundamental frequency median and range per talker per condition. Two participants were excluded from the fundamental frequency measures because of their frequent use of creaky voice, which lead to unreliable measures being obtained from their speech.

2.2.3.3. Long-term average spectrum (LTAS) measures

Another Praat script was used to calculate the LTAS measure for all the normalised files. The non-normalised files were not included in this analysis. The script removed the silences from each file, and calculated LTAS by acquiring the values for the first 100 bins using a 50 Hz bandwidth (0-5000 Hz). Then, the mean energy between 1 and 3k Hz was obtained by calculating the mean of the values between both frequencies.

2.2.3.4. Mean word duration measures

As a measure of speech rate, the average duration of the words in the files was calculated using a series of scripts[6]. First, a Praat script was used to calculate the duration of each annotated region in the TextGrids. Each region was then tagged as being either speech (SP), agreement (AGR), breath (BR), filler (FIL), garbage (GA), hesitation (HES), laughter (LG), or silence (SIL). The measure of mean word duration was then obtained by calculating the duration of all the SP tokens and dividing it by the total number of tokens tagged as SP. Finally, the measures were averaged across both diapix tasks to obtain a single measure of mean word duration for each talker per condition. On average, the duration of 1540 words in the Finnish NB condition, 1390 words in the Finnish VOC condition, 1158 words in the English NB condition and 1502 words in the English VOC condition were measured to obtain the mean word duration per participant.

[6] Mean word duration was used as a measure of speech rate to ensure consistency with previous studies using similar methods (e.g., Hazan & Baker, 2011).

2.3. Segmental measures

2.3.1. Sentence reading materials

The segmental measures were carried out on sentences designed to elicit specific segmental contrasts in two phonetic categories: bilabial plosives and vowels.

2.3.1.1. Bilabial plosives

To explore whether speakers modify their VOT differently in clear speech in their two languages, the VOT of bilabial plosives was investigated. These segments were chosen due to their differences in English and Finnish phonology; the Finnish phoneme inventory only includes a short-lag stop category /p/ while English contrasts short-lag /b/ and long-lag /p/ stops. If segmental clear speech modifications depend on the phonological contrasts in the language and are therefore language-specific, we would predict that although the Finnish and English short-lag stops are similar in casual speech, in clear speech the English short-lag /b/ would be decreased to make it as distinct as possible from the long-lag stop. The Finnish short-lag stop, on the other hand, would not change in VOT because it does not have to be made more distinct in VOT from any other segment.

English and Finnish sentences containing VOT keywords were created (see Appendix B); sentences with four minimal pairs of English keywords with initial /p-b/ were used, a subset of those used in the recordings of the LUCID database (Hazan & Baker, 2011), with some minor adaptations. Four Finnish keywords containing the voiceless /p/ were matched with the four minimal pairs of English keywords for following segments and initial syllable structure. The Finnish VOT sentences were created to match the English VOT sentences for number of syllables and for key syllable position in the sentence (either second or next to last). Phrasal stress was also matched as closely as possible, with the sentences having very similar if not identical phrasal stress patterns cross-linguistically. Additionally, the keywords were positioned in the sentence before a word ending in a vowel. Altogether, the sentences were constructed so that each keyword was produced twice in the same condition in English, and four times in Finnish, leading to 16 VOT sentences per language.

2.3.1.2. Vowels

The primary cue for the tense-lax distinction in English is spectral, but the durational distinction is more important in Finnish (Ylinen et al., 2010). This leads to a prediction that, in clear speech, speakers should enhance the spectral distinctions between English vowels to a greater extent than between Finnish vowels, but that they would enhance temporal distinctions between Finnish vowels more than for English vowels (Wassink et al., 2007; Kang & Guion, 2008). The high front vowel minimal pair /i-ɪ/ (tense-lax) in English and the long-short distinction /iː-i/ in Finnish were used in this investigation. Eight minimal pairs of keywords containing these contrasts were constructed, four for each language (see Appendix B). Cross-linguistically the vowels were matched for segmental context, with each vowel occurring before or after /s/, nasals or stops. As the short-long vowel distinction rarely occurs in closed syllables in Finnish, the vowels were not elicited in words of similar syllable structure cross-linguistically: the English vowels were placed in closed syllables, while the Finnish ones were in open syllables. Similarly, because of a lack of monosyllabic words in Finnish, the Finnish vowels were elicited in disyllabic words. The sentences in which the keywords were elicited had a similar number of syllables per sentence (English: 9 syllables, Finnish: 11 syllables) and

contained similar phrasal stress patterns. Additionally, all the keywords were placed in quotation marks so that all the words would be treated similarly, even though some of them differed in word category. All vowel sentences had a similar sentence structure, with the English sentences consisting of "[Person's name] [verb] that [KEYWORD] is the [adjective] [noun]" and the Finnish ones structured as "[Verb-1st or 2nd person] että [KEYWORD] on [adjective/noun] [noun]". Therefore all keywords were situated between "että" (*that*) and "on" (*is*) in Finnish, and "that" and "is" in English, which aimed at reducing any cross-linguistic differences in vowel enhancement due to different prosodic phrasing. Each keyword was placed in two different sentences, with 16 vowel sentences elicited per condition per language. Appendix C displays example sentences from the reading task.

2.3.2. Procedure

In each recording session, each participant completed the sentence reading task in the same language as the diapix recordings in that session; recordings were made in an acoustic booth, with sentences presented on a screen. In each pair, one participant did the sentence reading task before the diapix tasks, and the other participant completed it afterwards. The 32 sentences in each language were pseudo-randomised so that the same keyword did not occur two sentences in a row. DMDX (Forster & Forster, 2003) was used to present and record the sentences, using a sample rate of 22,050 Hz. For the casual condition, the subjects were asked to say the sentences "as if talking to a friend", and for the clear condition the instructions were to say the sentences "as if talking to someone with a hearing impairment". Each participant read 32 sentences in the casual condition, and 32 sentences in the clear condition for each language. Therefore 128 sentences were recorded per participant: 32 Finnish /p/ and 16 each of English /p/, English /b/, Finnish /iː/, Finnish /i/, English /i/ and English /ɪ/ sentences. A total of 1408 sentences were used in the study.

2.3.3. Processing

2.3.3.1. VOT measures

A Praat TextGrid was created for each of the VOT sentences produced in the sentence reading task. VOT was segmented manually using a 10 ms window as the interval between the stop burst and the zero crossing of the first glottal cycle of the vowel[7]. A few sentences in the casual condition were excluded due to incomplete closure of the vocal tract during stop production. 688 VOT sentences were used altogether, with an average of 7.8 English /b/, 7.7 English /p/ and 15.7 Finnish /p/ tokens used per speaker per condition. The duration of each interval was determined using a Praat script, and the mean VOT for each participant in each condition was calculated.

2.3.3.2. Temporal-spectral vowel measures

In Praat TextGrids, vowels were segmented manually in a 10 ms window from the zero crossing of the first glottal cycle of the vowel to the zero crossing of the offset of its last glottal cycle. A few vowels were deemed unsegmentable due to vowel nasalisation or unclear consonant boundaries and, as a result, a total of 695 vowel sentences were used in the analysis (mean: 7.9 Finnish /iː/, 7.9 Finnish /i/, 7.8 English /i/ and 8 English /ɪ/ tokens used per speaker per condition). The duration of

[7] As the word-initial stop consonants were preceded by a vowel in all of the sentence contexts, the voicing in the sequences was often coarticulated. Therefore it was impossible to measure whether prevoicing occurred in the stops.

each vowel interval was calculated automatically. The F1 and F2 values of the midpoint of the vowel were obtained using a Praat script. The means for each measure for each participant were used in the analysis.

3. Results

3.1. The difficulty of the diapix vocoder (VOC) condition compared to the "no barrier" (NB) condition

We must first establish whether the VOC condition was successful in increasing communication difficulty. This was examined using transaction time, i.e. the time taken to find the first eight differences in the picture. Transaction time has been found to be a reliable measure of communication efficiency, and to be longer in conditions in which communication was impaired either due to an adverse listening condition (Hazan & Baker, 2011) or due to a mismatch in native languages between participants (Van Engen et al., 2010).

Table 2. Mean transaction times in seconds for the diapix task in the NB and VOC conditions in Finnish (L1 Fin) and English (L2 Eng), with standard deviations in brackets.

	L1 Fin	s.d.	L2 Eng	s.d.
NB picture 1	325	(146)	298	(43)
NB picture 2	284	(90)	233	(37)
NB mean	305	(110)	266	(37)
VOC picture 1	343	(135)	404	(102)
VOC picture 2	277	(61)	350	(98)
VOC mean	310	(91)	377	(99)

A repeated measures ANOVA was run on the mean transaction times for each pair of Finnish-English talkers with within-subject factors of language (Finnish, English) and condition (NB, VOC). Only the interaction between language and condition was significant [$F(1,5)=9.014$; $p<0.05$; df=5]. Bonferroni-corrected paired t-tests revealed that the transaction time measures were significantly longer for the English VOC condition than the English NB condition [$t=-3.362$; $p<0.05$, df=5], suggesting an increase in communication difficulty. However, the transaction times for the Finnish VOC condition were not significantly longer than those for the Finnish NB condition [$t=-0.205$; n.s.; df=5]. A possible reason for the lack of increased transaction time in Finnish is the individual strategies used by the pairs in approaching the task. Two pairs who started with the diapix task in Finnish were very detailed in their description of the first Finnish NB picture, leading to higher overall mean transaction times in the NB condition than in the VOC condition for those pairs. Their transaction times for the second NB picture and for both VOC pictures were, however, within a more normal range. Excluding these two pairs seems to result in higher mean transaction times for the Finnish VOC condition (291 s) than the NB condition (252 s). Inspection of the clear speech strategies of the two pairs also shows no obvious differences as compared to the other eight participants. This suggests that, as with the other pairs, communication was more difficult for them in the VOC condition than in the NB condition in Finnish.

3.2. Stability of the measures within condition in the diapix task

Consistency within conditions in the diapix task is important for measures to be reliably compared across conditions. As two diapix pictures were completed per condition, and the lexical content of the speech in such spontaneous speech tasks may vary, within-condition consistency was checked by running a repeated measures ANOVA on language (Finnish, English), picture number (1^{st}, 2^{nd}) and condition (NB, VOC) on mean word duration, f0 range and median f0. Measures of mean energy were not used due to the small number of participants for whom data from both pictures were available. There was neither a main effect of picture nor an interaction of condition and picture for any of the measures examined. These results suggest that the measures are stable within-condition. Therefore across-condition comparisons can be attributed to real differences between conditions, rather than to the inherent variability present within-conditions in the diapix task.

3.3. The clear speech strategies used by the bilinguals in Finnish and English

3.3.1. Global measures – the diapix task

To examine whether the Finnish-English late bilinguals use similar global enhancement strategies in their two languages, the speakers' clear speech production in both their languages was compared. As Finnish is typologically very different to English, the speakers may enhance different aspects of the signal in the two languages. This would imply that language-dependent enhancement applies not only for segmental but for global measures too. On the other hand, if the global modifications only serve to enhance the overall salience of the signal, then speakers should use similar strategies across languages.

Repeated measures ANOVAs were carried out on each of the measures of median f0, f0 range, mean energy 1-3 kHz and mean word duration with language (L1 Finnish, L2 English) and condition (NB, VOC) as within-subject factors.

3.3.1.1. Median f0 and f0 range

Median f0 did not vary across languages. There was a main effect of condition [$F(1,9)=6.328$; $p<0.05$]: when speaking in the VOC condition, there was a slight increase in median f0 as compared to the NB condition (198 Hz vs. 189 Hz). However, there was no interaction between the two factors and therefore the speakers used this same strategy across their languages. Although there was a near-significant main effect of language for f0 range [$F(1,9)=5.02$; $p=0.052$], with a wider f0 range for Finnish than for English speech (45.8Hz vs. 39.1Hz), it was probably due to the more frequent use of creaky voice in Finnish than in English casual speech. There was no difference in the strategies used for f0 range, however; the interaction between language and condition was not significant. The result for f0 range is as expected: previous research which has elicited clear speech using a vocoder has also found minimal differences in f0 range as compared to the baseline condition (Hazan & Baker, 2011); here, we found that f0 range was not varied according to speaking style in either English or Finnish.

Table 3. Means and standard deviations (in brackets) of median f0, f0 range, mean energy (ME) 1-3k Hz and mean word duration (MWD) in the "casual" diapix_NB and "clear" diapix_VOC conditions for bilinguals (L1 Finnish, L2 English) and native speakers (N English). The measure of percent change between the VOC and NB conditions is also included.

	Diapix_NB			diapix_VOC			% change		
	L1 Finnish	L2 English	N English	L1 Finnish	L2 English	N English	L1 Finnish	L2 English	N English
f0 median (Hz), n=10	188.4 (15.6)	188.7 (16.3)	199.2 (12.0)	196.4 (15.6)	199.8 (13.6)	210.7 (14.0)	4.5 (6.4)	6.4 (8.7)	5.8 (4.8)
f0 range (Hz), n=10	48.2 (23.9)	37.8 (9.8)	37.0 (5.6)	43.4 (9.7)	40.5 (10.3)	42.7 (10.0)	1.4 (31.5)	10.8 (27.6)	16.0 (24.0)
ME 1-3kHz (dB), n=6	26.1 (2.1)	26.5 (1.4)	26.7 (2.0)	28.7 (1.6)	29.0 (1.2)	28.1 (2.6)	9.0 (3.7)	9.5 (3.5)	5.2 (5.7)
MWD (ms), n=12	315.6 (23.1)	307.9 (32.5)	254.8 (25.8)	395.9 (53.8)	369.3 (44.7)	338.3 (48.0)	26.0 (19.2)	20.6 (15.0)	33.4 (18.6)

3.3.1.2. Mean energy (ME 1-3k Hz)

There was no effect of language, but a highly significant main effect of condition [$F(1,5)=39.194$; $p=0.001$]. In the VOC condition, the participants' speech had a higher mean energy than in the NB condition (29 dB vs. 26 dB). The interaction was, again, not significant, and therefore the speakers used similar strategies in increasing the intensity of their speech in both languages when speaking clearly.

3.3.1.3. Mean word duration

The interaction between language and condition was significant [$F(1,11)=6.821$; $p<0.05$]; paired t-tests show that mean word duration increased in clear speech both in Finnish (VOC: 396ms; NB: 316ms) [$t=4.806$; $p<0.001$; $df=11$] and in English (VOC: 369ms; NB: 308ms) [$t=4.757$; $p<0.001$; $df=11$]. A further paired t-test on the percent change in mean word duration between conditions in the two languages shows a trend towards Finnish words being lengthened more than English words (Finnish: 26%; English: 21%) [$t=2.083$; $p=0.061$; $df=11$]. Therefore, although the speakers decrease their speech rate in both Finnish and English, there is a trend for a larger decrease in their Finnish than in their English clear speech. However, the comparison of speech rate across the two languages is fairly unreliable due to the many structural differences in words between the two languages. For example, Finnish maintains a rich inflectional morphology, contrary to English. Therefore it is likely that Finnish words are longer than English words, which is likely to impact on mean word duration.

Summary: Overall, of the four global measures, only changes in speech rate differed in magnitude between the talkers' two languages. Importantly, these results demonstrate that despite the many differences between the two languages, Finnish-English late bilinguals seem to use the same global enhancement strategies involving fundamental frequency and mean energy in both languages. Table

3 summarises the data for each language and condition, and table 4 presents a summary of the statistical results.

Table 4. Repeated measures ANOVA on language (L1 Finnish, L2 English) and condition (diapix_NB, diapix_VOC) for each measure. The statistically significant results are marked with an asterisk (: p≤0.05; **: p≤0.01).*

		F	df	p	
median f0	language	0.62	(1,9)	0.450	
	condition	6.33	(1,9)	<0.05	*
	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)	<0.05	*
	condition	23.93	(1,11)	<0.001	**
	interaction	6.82	(1,11)	<0.05	*

3.3.1.4. Individuals' use of the global enhancement strategies

To examine whether similar enhancement strategies are used in the two languages also at the individual level, correlation analyses were run on the percent change in measures in the two languages (% change in Finnish vs. % change in English)[8]. Significant positive correlations were found for mean word duration [r=0.890; p<0.001; R^2=0.79; n=12] and f0 range [r=0.853; p<0.01; R^2=0.72; n=10], and there was a trend for changes in f0 median to be correlated across languages [r=0.566; p=0.088; R^2=0.32; n=10]. The results indicate that individual speakers were using similar global strategies in speaking clearly in both their languages.

3.3.2. Segmental measures – the sentence reading task

3.3.2.1. Bilabial plosives

VOT measures were examined to compare the speakers' clear speech enhancement strategies for the two short-lag stops, English /b/ and Finnish /p/. If segmental clear speech modifications are language-specific, and made to enhance phonological contrasts in a language, we would expect the VOT for English /b/ to decrease in clear speech to distinguish it from the long-lag English /p/. We would not expect the Finnish short-lag stop to be modified in clear speech.

[8] The analysis was not run on the measure of mean energy due to the small number of participants for whom mean energy measures were available.

Table 5. Mean VOT values (in ms) for the Finnish /p/ and English /b/ in casual and clear conditions. The t-test columns and rows indicate the t- and p- values of paired t-tests run on adjacent cells. Df=10 for each test. The statistically significant results are marked with an asterisk (: p<0.05; **: p<0.01).*

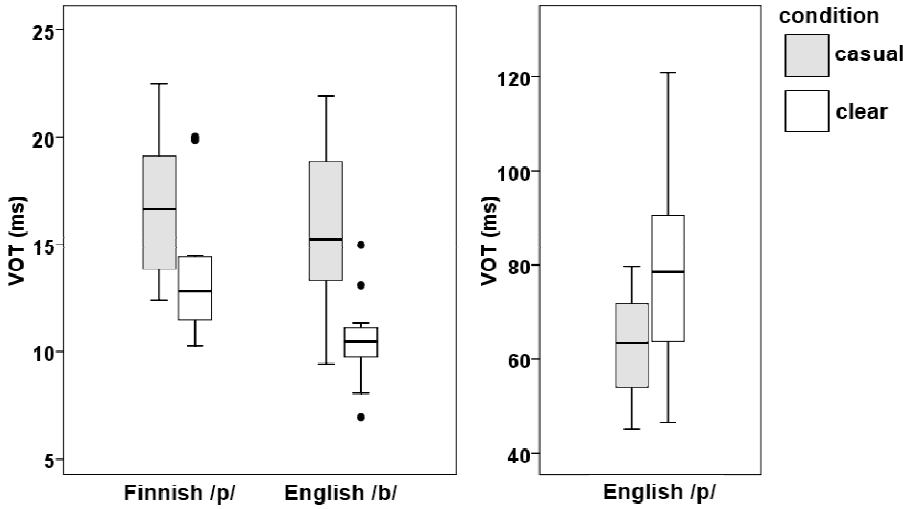
	casual	clear	t-test	p
Finnish /p/	16.7	13.7	t=3.522	<0.01**
English /b/	15.8	10.6	t=4.091	<0.01**
t-test	t=-0.724	t=-4.229		
p	0.485	<0.01**		

A two-way ANOVA was run on the VOT durations for English /b/ and Finnish /p/ with within-subjects factors of segment (/b/, /p/) and condition (casual, clear). The main effect of condition was significant [$F(1,10)=10.706$; $p<0.01$] but is modulated by a significant interaction between segment and condition [$F(1,10)=5.861$; $p<0.05$], suggesting a difference in clear speech strategies in the two languages. Table 5 displays the results of Bonferroni-corrected paired t-tests. Although the VOT for both segments is significantly decreased in both languages [Finnish: $t=3.522$; $p<0.01$; $df=10$; English: $t=4.091$; $p<0.01$; $df=10$], the VOT for Finnish /p/ is decreased less (3.0 ms) than that of English /b/ (5.2 ms) [$t=-2.421$; $p<0.05$; $df=10$](see figure 1). Although this cross-language difference is small, the results indicate that speakers modify both segments, but suggests that speakers are attempting to make the phonological categories in English as distinct as possible when speaking clearly.

To explore whether the speakers also enhance the VOT of English /p/, a paired samples t-test was performed on the English /p/ for casual and clear conditions. The VOT in the clear condition was longer (79 ms) than in the casual condition (63 ms) [$t=-2.390$; $p<0.05$; $df=10$] (see figure 1).The bilinguals are enhancing the VOT of the long-lag English stop in the opposite direction to the short-lag stop, further confirming the expectation that the two English stops are being made as distinct from each other as possible in clear speech. Note, though that this increase in VOT could partly be due to a slower speech rate in the clear speech condition.

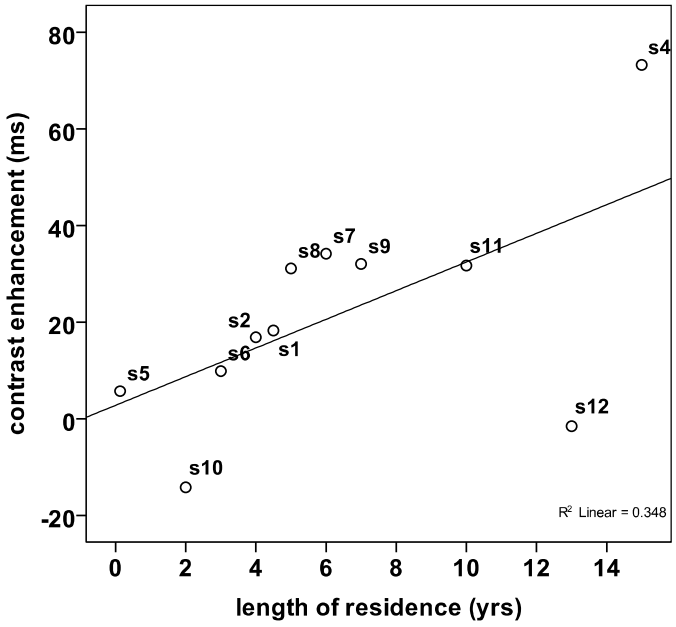
A correlation analysis was run on the amount of contrast enhancement (i.e., difference between the VOT in the English /p/ and /b/ in casual and clear speech for each speaker) and the amount of experience in the language by the speaker (length of residence in an English-speaking country). There is a trend towards a greater enhancement of the contrast for those speakers who have resided in the UK for longer [$r=.590$; $p=0.056$; $R^2=0.348$] (see figure 2). Although the result is preliminary, it supports the proposal that contrast enhancement is language-dependent and reflects the amount of experience in a language.

Figure 1. VOT values (in ms) of short-lag Finnish /p/ and English /b/ and long-lag English /p/ in the casual conditions (grey boxes) and clear conditions (white boxes). The dots mark any outliers.



In summary, the results reveal that the speakers modify the two short-lag stops, English /b/ and Finnish /p/, differently in clear speech; the VOT for English /b/ is decreased to a greater extent than the Finnish /p/, and the VOT for English long-lag /p/ is increased in clear speech. This indicates that phonological contrasts are enhanced in clear speech, while segments which do not maintain phonological contrasts on a certain dimension do not need to be modified to as great an extent. There seems to be a greater amount of contrast enhancement between the English /b/ and /p/ for speakers with greater experience in the language. These results support the hypothesis of language-experience-dependent enhancement of segmental contrasts.

Figure 2. The relationship between the amount of contrast enhancement of the English /b-p/ distinction for each speaker, and the speakers' length of residence in an English-speaking country.



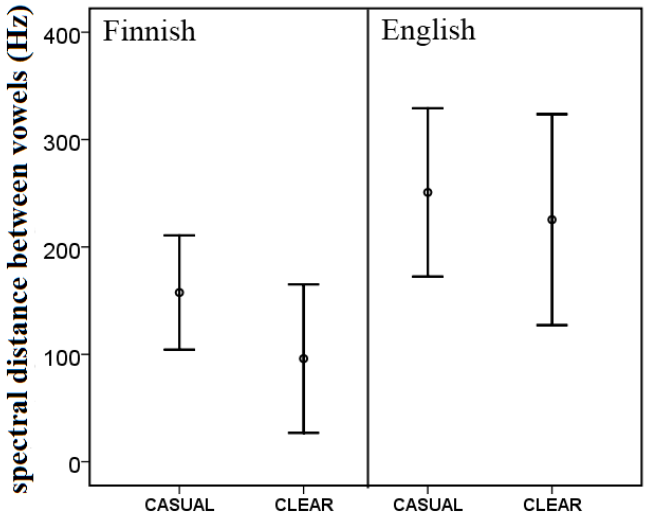
3.3.2.2. Vowels

The speakers' strategies for enhancing the high front vowels in the English tense/lax and Finnish long/short contrasts were also examined. The primary cue for the English vowel contrasts is spectral whereas vowel duration is the primary cue for the Finnish contrast. Therefore a larger amount of spectral enhancement is expected for the English vowels as compared to the Finnish vowels, but the speakers were expected to enhance the temporal aspect to a greater extent in Finnish than in English.

For each participant, the F1/F2 Euclidian distance was calculated from the mean F1 and F2 values for each vowel per condition, by taking the square root of the sum of squares for F1 and F2. Then, as a measure of the spectral distance between the two high front vowels in the two languages in each condition, the Euclidean distance values for the short/lax vowels were subtracted from those of the long/tense vowels for each speaker in each condition. An ANOVA was run on the measure with within-subjects factors of language (Finnish, English) and condition (casual, clear). If speakers enhance the spectral distance between the two vowels to a greater extent in English than in Finnish, the interaction of language and condition is expected to be significant.

The effect of language was significant; the spectral distance between the Finnish high front vowels was smaller than in the English high front vowels (126.8 Hz vs. 238.1 Hz) [$F(1,10)=9.683$; $p=0.011$]. There was also an effect of condition: surprisingly, the spectral distance between the high front vowels was smaller in the clear condition than in the casual condition (160.7 Hz vs. 204.2 Hz) [$F(1,10)=5.901$; $p<0.05$] (see figure 3). However, there was no interaction of language and condition [$F(1,10)=1.696$; n.s.], and therefore the speakers treated spectral distance similarly in both languages.

Figure 3. 95% confidence intervals of spectral distances between the two high front vowels in Finnish and English, in the casual and clear conditions across the 11 speakers.

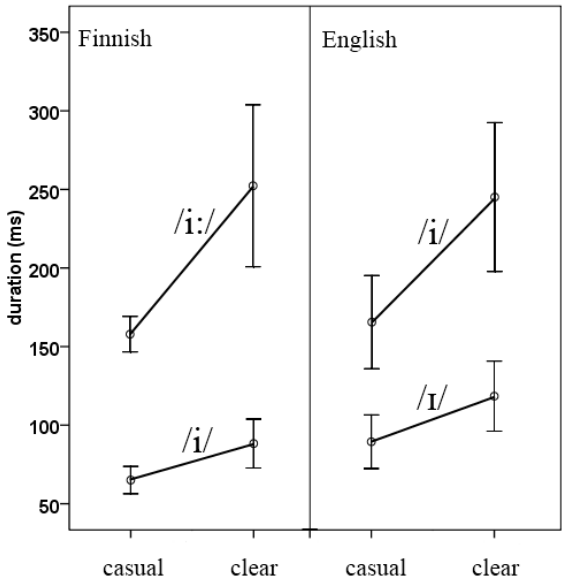


To examine whether the speakers enhance the temporal ratios between the short and long vowels to a greater extent in Finnish than in English, the temporal ratios for each language in each condition

for each participant was calculated (long or tense vowels divided by short or lax vowels). An ANOVA was run on the measure, with within-subject factors of language (Finnish, English) and condition (casual, clear). There were higher ratios between the short and long vowels in Finnish (2.79) than in English (2.05) [$F(1,10)=24.33$; $p=0.001$]. The ratios were also greater in clear speech (2.59) than in casual speech (2.25) [$F(1,10)=28.87$; $p<0.001$]. However, the interaction between language and condition was not significant, and therefore the speakers were found to use the same strategies of enhancing the vowel duration ratios in both their languages. An ANOVA was also run on the difference in duration between the long/tense and short/lax vowels in each language, but the results were the same as for the vowel ratios; there were main effects of language and condition, but no interaction between language and condition [$F(1,10)=2.925$; n.s.] (see figure 4).

Altogether, the hypothesis of a greater enhancement of the spectral distance between vowels in English than in Finnish, and a greater temporal enhancement in Finnish, does not hold. However, the significant main effects of language for both the spectral and durational measures show that the speakers generally produce the vowels in the two languages differently. There was a greater spectral distance between the English high front vowels than the Finnish high front vowels, and a greater durational distance between the Finnish than the English high front vowels. The results suggest that although speakers may use different cues for producing the vowels in the different languages, in clear speech both spectral and durational enhancement is applied to a similar extent.

Figure 4. Durations for the bilinguals' English tense/lax and Finnish short/long vowels in the casual and clear conditions (error bars: +/- SD).



3.4. Comparison of English global clear speech modifications of Finnish-English bilinguals and native speakers

Next, we investigated whether the Finnish-English bilinguals' clear speech strategies approximate those of native English speakers.

As a comparison to the L2 English of the Finnish-English bilinguals, the acoustic-phonetic measures from diapix recordings of native English speakers were used from the LUCID database (Hazan & Baker, 2011). They were native Southern British English speakers, with an age range of 19 to 29 years, and the first six female pairs were used. For the measures in which some of the Finnish-English participants had to be excluded, the native English participant with the same participant number was also excluded. As the native speakers completed three diapix pictures in both the NB and VOC conditions, but the bilinguals completed only two pictures per condition, the native speaker recordings from only the first two diapix pictures were used in the current comparison.

As a more reliable comparison across speakers, due to the inherent variability of speakers' baseline values in their casual speech, the percent relative change from the NB to the VOC condition for each of the median f0, f0 range, mean energy 1-3k Hz and mean word duration measures was calculated. An independent samples t-test was then run on each measure to investigate the effect of speaker group (bilingual, monolingual).

There were no significant differences in the extent of clear speech modifications made for median f0, f0 range, mean energy or mean word duration between non-native and native English speakers (see the means and standard deviations of each measure in table 3). Altogether, the results suggest that these highly-proficient late bilinguals were able to modify their speech to accommodate their interlocutors similarly to native speakers.

4. Discussion

This study assessed the extent to which different acoustic-phonetic modifications in clear speech are language-dependent or language-independent by investigating the clear speech strategies of Finnish-English late bilinguals in both their languages. Specifically, the experiment tested the hypothesis that global enhancements increase the overall salience of speech, and are therefore language-independent, while language-dependent segmental modifications enhance the phonological contrasts between categories (Bradlow & Bent, 2002). The findings of the study support this proposal to some extent: a comparison of the acoustic-phonetic adjustments made by the speakers in both their languages in the "vocoded" diapix_VOC and "no-barrier" diapix_NB conditions revealed that the speakers use similar global clear speech strategies in Finnish and English, despite several typological differences between the languages. The exception was speech rate, measured as mean word duration, as speakers were found to lengthen their words in clear speech in Finnish more than in English. However, it is possible that these differences arose from the structural differences between Finnish and English words; this interpretation is further supported by comparisons between the speakers' L2 English and native English speakers' speech rate, which showed no differences in enhancement strategies, and by findings showing that the relative change in speech rate in Finnish was correlated with that of English for the speakers. The result agrees with previous findings indicating similar enhancement strategies across languages for f0 and intensity modifications (Smiljanic & Bradlow, 2005; Wassink et al., 2007). Although global changes may be influenced by language structure (Cho et al, 2011; Smiljanic & Bradlow, 2009a), and our study examined only a subset of possible global measures, our findings suggest that there may be a

greater tendency for global modifications to clear speech to be language-independent. This implies that these changes may be made to enhance the general auditory-perceptual salience of the signal.

As predicted by the hypothesis, the only language-specific strategies implemented by the speakers in the current experiment were on the segmental level: in the sentence reading task, carried out in both casual and clear speaking styles, enhancement strategies for VOT differed across languages. Although the Finnish short-lag /p/ and English short-lag /b/ have very similar VOTs in the casual condition, the VOT of the Finnish /p/ decreased less in clear speech than the English /b/. The greater VOT decrease for /b/ may be guided by the principle of contrast enhancement: the speakers are attempting to increase the acoustic distance between the "voiced" and "voiceless" categories in English. However, although the "voiceless" Finnish /p/ has no "voiced" equivalent, its VOT is still lowered in clear speech, implying that segmental modifications may not necessarily always reflect contrast enhancement either: this may be due to increased tension in the vocal folds when speech is produced with more effort in clear speech. Alternatively, although in this study it was necessary to use bilinguals to avoid between-speaker effects, the bilinguals may have been influenced by the VOT of the English stops even in their native language (Flege, 1987). It is particularly interesting that the VOT is decreased for the short-lag stops when the speech rate decrease in clear speech would imply longer durations for all segments. Further supporting the hypothesis for contrast enhancement, speakers also increased the VOT of the English long-lag /p/, a strategy found for native English speakers (Smiljanic & Bradlow, 2008). Therefore not only are the speakers using different clear speech strategies for the "similar" short-lag English /b/-Finnish /p/ categories, but they are also enhancing the "new" aspirated /p/ category by modifying VOT in the opposite direction. This contrasts with the findings of Smiljanic & Bradlow (2009b) for Croatian speakers of English whose length of residence in an English-speaking country was similar to the participants in the current experiment. The Croatian speakers were found to transfer their L1 clear speech strategy of applying prevoicing to the English /b/, but they did not lengthen the VOT of English /p/. It is possible that the existence of the /p-b/ distinction in Croatian but not in Finnish makes the application of more native-like strategies more difficult for the Croatian L2 speakers.

The language-specificity of clear speech strategies was not evident in the vowel data. As in Wassink et al. (2007), speakers used the same strategies for enhancing the spectral and temporal aspects of the vowels in the two languages, despite Finnish and English weighting those cues differently, and despite the speakers having spectrally more distinct English vowels and temporally more distinct Finnish vowels. This contradicts previous findings of different enhancement strategies for Korean stops according to the speakers' use of the primary cue (Kang & Guion, 2008). It is possible that this disparity can be explained by different clear speech principles operating on consonants than on vowels: as discussed in the introduction, an important finding of clear speech research has been that vowels are spectrally enhanced regardless of their confusability with other vowels (Bradlow, 2002; Smiljanic & Bradlow, 2005). These enhancements may instead be language-independent; the decreased speech rate in clear speech may allow speakers to approximate vowel targets and therefore enhance vowel cues regardless of the specific cues used in the language. This implies that the H&H theory (Lindblom, 1990) may not hold for vowels: minimal effort may not always be spared (Bradlow, 2002; Cho et al., 2011). On the other hand, the Finnish-English late bilinguals may be transferring their clear speech strategies for vowel enhancement from their L1 to their L2 because of the similarities between the Finnish /iː-i/ and English /i-ɪ/ vowel pairs. Therefore it may be that the

late bilinguals simply do not have different primary vowel cues for the two languages, as previous studies on less proficient Finnish speakers of English have found (Ylinen et al., 2010). Further research into the ways in which the late bilinguals' Finnish and English vowel enhancement strategies differ from native English speakers' or early bilinguals' could elucidate the matter. Additionally, measures of vowel space expansion would enable a fuller assessment of the differences between the speakers' enhancement strategies for Finnish and English vowels.

The experiment also investigated whether the late bilinguals were able to modify their speech to the listener in their second language similarly to native English speakers. Contrasting the speech produced by the 12 late bilinguals and 12 native speakers of English in the "vocoded" diapix_VOC and "no barrier" diapix_NB conditions revealed that the late bilinguals were able to produce global acoustic-phonetic modifications in their second language to a similar extent as native English speakers, and they were able to do this without specific instruction. Using the native speaker data from the LUCID corpus, Hazan & Baker (2011) found that speakers tailored their clear speech according to the communication barrier, with greater changes made to f0 median and range when speaking to someone hearing their speech masked by babble noise rather than vocoded. The Finnish-English late bilinguals in the present study also did not make large changes to their fundamental frequency when their speech was vocoded. This implies that the late bilinguals can modify their speech in their L2 to suit the specific needs of the listener. Therefore the H&H theory (Lindblom, 1990) seems to extend to late bilingual speakers too. These kinds of speech modifications may be very important in everyday communication as they probably promote speaker intelligibility in various different contexts (Smiljanic & Bradlow, 2009a).

The results from the current study generally support the proposal that previous findings of perceptually less beneficial clear speech produced by non-native speakers (Rogers et al., 2010; Li, 2009) and a smaller intelligibility gain in the perception of clear speech by non-native listeners (Bradlow & Bent, 2002; Bradlow & Alexander, 2008) reflect the inexperience of the speakers and listeners in their non-native language (Smiljanic & Bradlow, 2009a). Non-native listeners are probably able to make use of the global clear speech modifications in their second language, and are able to produce them, because the clear speech adjustments are similar to those used in their native language. On the other hand, they can probably make less use of and apply the enhancement of segmental detail in their L2; although the highly proficient Finnish-English late bilinguals of this study are able to enhance some of the non-native phonetic detail in clear speech, the relationship between experience with the language and the amount of contrast enhancement for the English /b-p/ distinction tended towards significance. A closer examination of the clear speech strategies of less proficient L2 speakers and the intelligibility gain in both the bilinguals' languages is needed to further assess the issue.

A shortcoming of the current study is that spontaneous speech was used to measure global modifications, while read speech was used for the segmental detail. Although this method was necessary to control for the environment in which the segmental detail was elicited across languages, it is possible that global and segmental changes elicited from the same speech material would have yielded different results. Further work is ongoing using a spontaneous speech task to elicit keywords in a communicative context.

Acknowledgements

We would like to thank Ocke-Schwen Bohn and three anonymous reviewers for helpful comments on an earlier version of this paper. We are grateful to Martti Vainio and colleagues at Helsinki University and Aalto University for carrying out the orthographic alignment for the Finnish files. We also thank Ann Bradlow and colleagues from the Dept of Linguistics at Northwestern University for making transcription and alignment software available to us and for fruitful discussions about this work. This project was partially funded by the UK Economics and Social Research Council (RES-062-23-0681).

Appendix A. Participant information.

Further information on the participants, with the age they started to learn English at school, their number of years of learning English formally, their age on moving to the UK, length of residence (LoR) and % language use.

subject number	age	age Eng at school	years of learning	UK move age	LoR (yrs)	% use Finnish	% use English
1	25	8	10	20	4.5	20	80
2	24	9	10	19	4	35	60
3	26	8	18	N/A	0	80	18
4	34	8	10	18	15	10	90
5	27	8	11	27	0.125	20	80
6	32	12	10	26	3	2	98
7	30	11	11	23	6	30	70
8	31	9	11	26	5	40	60
9	29	10	10	22	7	50	50
10	20	10	9	16	2	30	70
11	35	9	10	25	10	25	75
12	31	9	12	18	13	1	99

Appendix B. The VOT and vowel keywords.

The words used in the sentence reading task to elicit VOT and the tense-lax or short-long distinctions. The syllable boundaries in Finnish are identified with the dash within-words. Some of the English vowel keywords were those used in Ylinen et al. (2010).

	context	English	Finnish	translation
1. VOT	/_i□z/, /_i□s/	bees , peas	piis-pa	“bishop”
	/_□i/	buy , pie	pai-ta	“shirt”
	/_et/	bet , pet	Pet-ri	a name
	/_in/	bin/s , pin/s	pin-ni /pins-si	“pin” / “brooch”
2.vowels	/s_m/	seem , sim	sii-ma , si-ma	“fishing line”, “mead”
	/t_n/	teen , tin	Tii-na , ti-na	a name , “tin”
	/k_n/	keen , kin	Kii-na , ki-na	“China”, “argument”
	/s_k/	seek , sick	sii-ka , si-ka	“whitefish”, “pig”

Appendix C. Example sentences.

Example sentences from the reading task. For ease of identification, keywords are in bold, but in the experiment all words were in a normal font.

English	Finnish	translation
The **bin** was knocked on the floor.	Se **pinni** löytyi maasta.	That pin was found on the ground.
The man put some of the **bins** out.	Neuleisiin kuuluu kai **pinssi**.	Cardigans need to have a brooch.
The **pin** was hit by the ball.	Se **pinni** kuuluu hiuksiin.	That pin belongs in your hair.
The girl picked some of the **pins** up.	Hiuksiini jäikin kai **pinssi**.	A brooch probably got left in my hair after all.
Freddie knows that '**sim**' is a new word.	Luulen, että '**sima**' on hyvä sana.	I think that 'sima' is a good word.
Terry thought that '**seem**' is the last word.	Tiedän, että '**siima**' on vanha sana.	I know that 'siima' is an old word.

References

Baker, R., & Hazan, V. (2011). DiapixUK: a task for the elicitation of spontaneous speech dialogs. *Behavior Research Methods,* 43 (3), 761-770.

Bent, T., Buchwald, A. & Pisoni, D. B. (2009). Perceptual adaptation and intelligibility of multiple talkers for two types of degraded speech. *Journal of the Acoustical Society of America, 126*, 2660-2669.

Boersma, P. & Weenink, D. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*, 341-345.

Bradlow, A.R. (2002). Confluent talker- and listener-related forces in clear speech production. In Gussenhoven, C. & Warner, N. (Eds.) *Laboratory phonology 7*, (pp. 241–73). Berlin /New York: Mouton de Gruyter.

Bradlow, A. R., & Alexander, J. (2007). Semantic-contextual and acoustic-phonetic enhancements for English sentence-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America, 121(4),* 2339–49.

Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America*, *112(1),* 272–84.

Bradlow, A. R., Kraus, N. & Hayes, E. (2003). Speaking clearly for learning-impaired children: sentence perception in noise. *Journal of Speech, Language, and Hearing Research, 46*, 80–97.

Charles-Luce, J. (1997). Cognitive Factors Involved in Preserving a Phonemic Contrast. *Language and Speech*, 40(3), 229.

Cho, T., Lee, Y., & Kim, S. (2011). Communicatively driven versus prosodically driven hyper-articulation in Korean. *Journal of Phonetics*, 39(3), 344-361.

Davis, M.H., Johnsrude, I.S., Hervais-Adelman, A.G., Taylor, K.J. & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology*, 134, 222-241.

Ferguson, S. H., & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America, 112*, 259–71.

Ferguson, S. H. & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research, 50*, 1241–55.

Flege, J.E. (1987). The production of 'new' and 'similar' phones in a foreign language: evidence for the effect of Equivalence Classification. *Journal of Phonetics,* 15, 47-65.

Forster, K. & Forster, J. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavioural Research Methods*, 35, 116-124.

Gut, U. (2006). Unstressed vowels in non-native German. *Speech prosody 2006.* Dresden.

Hansen, J.G. (2006). Acquiring a Non-Native Phonology. *Linguistic Constraints and Social Barriers*. London: Continuum.

Hazan, V. & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *Journal of the Acoustical Society of America*, 130 (4), 2139-2152.

Iivonen, A. (1998). Intonation in Finnish. In D. Hirst & A. Di Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp.311-327). Cambridge University Press.

Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language, 69*, 505–528.

Kang, K.H. & Guion, S.G. (2008). Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. *Journal of the Acoustical Society of America, 124*, 3909-3917.

Krause, J.C., Braida, L.D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *Journal of the Acoustical Society of America, 115*, 362-78.

Leather, K. & James, W. (1996). Second language speech. In W.C. Ritchie & T.K. Bhatia (Eds.), *Handbook of second language acquisition, (*pp.269-316). San Diego: Academic Press.

Li, C. (2009). *Perception of foreign-accented clear speech by younger and older English listeners*. Unpublished doctoral dissertation, Simon Fraser University.

Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling*, (pp.403–439). Amsterdam: Kluwer Academic.

Liu, S., Del Rio, E., Bradlow, A. R., and Zeng, F.-G. (2004). Clear speech perception in acoustic and electrical hearing. *Journal of the Acoustical Society of America, 116,* 2374-2383.

Moon, S.-J. & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *Journal of the Acoustical Society of America, 96*, 40- 55.

Munro, M., & Derwing, T. (1995). Processing time, accent, and comprehensibility in the perception of foreign-accented speech. *Language and Speech*, *38*, 289-306.

Munro, M., & Derwing, T. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *49*, Supplement 1, 285-310.

NUAligner, Northwestern University Linguistics Department software. http://groups.linguistics.northwestern.edu/documentation/nualigner_home.html Last accessed 04.02.11

Picheny, M. A., Durlach, N.I. & Braida, L.D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research, 28*, 96–103.

Picheny, M. A., Durlach, N.I., & Braida, L.D., (1986). Speaking clearly for the hard of hearing II: acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research, 29*, 434–46.

Ringen, C. & Suomi, K. (2009). Fenno-Swedish VOT: Influence from Finnish? *Proceedings, FONETIK 2009,* (pp.60-65). Dept. of Linguistics, Stockholm University.

Rogers, C.L., DeMasi, T. & Krause, J. (2010). Conversational and clear speech intelligibility of /bVd/ syllables produced by native and non-native English speakers. *Journal of the Acoustical Society of America, 128(1)*, 410-423.

Rosen, S., Faulkner, A. & Wilkinson, L. (1999). Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. *Journal of the Acoustical Society of America, 106(6)*, 3629-3636.

Sajavaara, K. & Dufva, R.H., (2001) Finnish-English Phonetics and Phonology. *International Journal of English Studies*, *1*, 241-256.

Scarborough, R., Brenier, J., Zhao, Y., Hall-Lew, L., & Dmitrieva, O. (2007). An Acoustic Study of Real and Imagined Foreigner-Directed Speech. *Proceedings of the XVIth International Congress of Phonetic Sciences*. Saarbrücken.

Smiljanic, R., & Bradlow, A.R. (2005). Production and perception of clear speech in Croatian and English. *Journal of the Acoustical Society of America*, *118(3)*,Pt. 1:1677–88.

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

Smiljanic, R. & Bradlow, A. R. (2009a). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Linguistics and Language Compass, 3*, 236–264.

Smiljanic, R. & Bradlow, A.R. (2009b). Native and non-native clear speech production. *Journal of the Acoustical Society of America*, *125(4),* Pt. 2, 2753.

Smiljanic, R., & Bradlow, A. R. (2011). Bidirectional clear speech perception benefit for native and high-proficiency non-native talkers and listeners: Intelligibility and accentedness. *Journal of the Acoustical Society of America*, 130(6), 4020-4031.

Uchanski, R. M. (1988). *Spectral and temporal contributions to speech clarity for hearing impaired listeners*. Unpublished Doctoral Dissertation. Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

Ulbrich, C. (2008). Acquisition of regional pitch patterns in L2. *Speech Prosody 2008*, (pp. 575-578). Campinas.

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.

Wassink, A., Wright, R. & Franklin, A. (2007). Intraspeaker variability in vowel production: an investigation of motherese, hyperspeech, and Lombard speech in Jamaican speakers. *Journal of Phonetics, 35*, 363–79.

Wavescroller, Northwestern University Linguistics Department software. http://groups.linguistics.northwestern.edu/documentation/wavescroller_home.html Last accessed 04.02.11

Ylinen, S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R., Näätänen, R. (2010). Training the Brain to Weight Speech Cues Differently: A Study of Finnish Second-language Users of English. *Journal of Cognitive Neuroscience*, 22(6), 1319-1332.