Acoustic Analysis of Vowels Spoken Clearly and Conversationally by Non-native English Speakers

Herman Chi Nin Li

School of Audiology and Speech Sciences
University of British Columbia,
British Columbia, Canada
herman@audiospeech.ubc.ca

Connie K. So

MARCS Auditory Laboratories
University of Western Sydney, Australia
c.so@uws.edu.au

Abstract

This study examines the acoustic properties of clear speech produced by non-native speakers of English and investigates whether the strategies that they employ to produce English vowels in clear speech are similar to those of native English speakers. Both groups of speakers exhibited similar acoustic-phonetic changes between clear and conversational speech.

1. Introduction

Clear speech is a speaking style adopted by talkers when they are instructed to speak clearly, as if talking in a noise background or communicating with hearing impaired individuals (Bradlow, Kraus, & Hayes, 2003; Ferguson & Kewley-Port, 2002; Payton, Uchanshki, & Braida, 1994; Picheny, Durlach, & Braida, 1985; Schum, 1996). It has been demonstrated that clear speech is more intelligible than conversational speech for various listener populations under different listening conditions. For instance, Picheny et al. (1985) found that hearing-impaired listeners benefit from clear speech more than conversational speech in quiet independent of the listener, presentation level, and frequency-gain characteristics. Clear speech is beneficial for normal-hearing and impaired-hearing listeners not only in noisy, but in reverberant environments (Payton et al., 1994). Also, it has been reported that clear speech produced by younger and older adults with normal hearing were more intelligible than conversation speech in noise for hearing-impaired older adults (Schum, 1996). The clear speech intelligibility advantage has also been found for cochlear implant adults (Iverson & Bradlow, 2002), for school-age children with and without learning disabilities (Bradlow et al., 2003), for non-native English listeners (Bradlow & Bent, 2002), and for non-English speaking subjects (Croatian) (Smiljanic & Bradlow, 2005). In addition to aforementioned studies that employed sentences as speech materials, clearly spoken English vowels have been found to benefit normal-hearing listeners in noisy conditions (Ferguson, 2004; Ferguson & Kewley-Port, 2002).

Previous research has established that intelligibility of clear speech is associated with a wide variety of acoustic-phonetic changes: an increase in intensity (Picheny, Durlach, & Braida, 1986), reduced speaking rate (Bradlow et al., 2003; Picheny et al., 1986), longer and more pauses (Krause & Braida, 2004; Picheny et al., 1986), an increase in fundamental frequency (F0) (Picheny et al., 1986), a wider F0 range (Bradlow et al., 2003; Picheny et al., 1986), an increase in vowel duration (Ferguson & Kewley-Port, 2002), a shift in the fist two formant frequencies (F1 and F2) (Ferguson & Kewley-Port, 2002), and an expanded vowel space (Smiljanic & Bradlow, 2005).

With the exception of a few research, most studies have focused on English clear speech. There is a paucity of data on clear speech production and/or perception for speakers of a second language. The goals of the present study are to examine (1) the acoustic properties of clear speech produced by non-native English speakers, and (2) whether the strategies that they employ to produce English vowels in clear speech are the same as those of native English speakers.

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after abstract only review

2. Methods

2.1. Speakers

Ten speakers of Hong Kong Cantonese (five female and five male) and a comparison group of native speakers of Canadian English (five female and five male) served as participants in this production study. All the Cantonese speakers of English were born and raised in Hong Kong. The mean age was 21 years ranging from 19 to 23 years. They all finished high school in Hong Kong before coming to study in Canada. Their mean length of stay in Canada was 29 months. The native English speakers were all born in western Canada. Their ages ranged from 18 to 30 with a mean of 24 years. All participants were undergraduate students. None of the speakers had any known voice and speech anomalies, and they all passed a pure-tone hearing screen prior to performing the production task.

2.2. Speech materials and procedures

For the current study, test words consisted of four English vowels (/i u æ a/) in /bVt/ context. The target words were embedded in a carrier sentence “Now I say bVt again”. Individual recordings were made using a head-worn microphone (MB Quart K800) in a sound-treated room. The speakers were recorded producing three instances of each of the sentences once in a conversational manner and once in a clear speaking style. In the first recording session, the native English speakers were instructed to read the sentences in a typical conversational manner, as though they were talking to their family members or friends. The Cantonese speakers of English were told to read the sentences conversationally, as if talking to native English speakers (e.g., friends, classmates, or cashiers in supermarkets) in their daily activities. In the second recording session, all speakers were instructed to read the sentences as if talking to an English-speaking older adult or hearing-impaired person. For both groups of speakers, all instructions were given in English.

2.3. Acoustic analysis

The sentences were randomized in each recording session. The recorded test words were excised and downsampled from 44010 Hz to 10 000 Hz for acoustic analysis. For the present study, vowel duration and values of fundamental frequency (F0) (mean, maximum, and minimum) were measured. Also, following Fergusion & Kewley-Port (2002), the first two formants (F1 and F2) at 50% of the vowel duration for each of the vowels produced by the speakers in each of the two speaking styles were measured using Colea Matlab code (Loizou, 1999) employing 12 LPC coefficient, a 20 msec hamming window, and a frame rate of 10 msec. Percentage change of pitch range (maximum F0 – minimum F0) were computed. Vowel space area based on the F1 and F2 values derived from each vowel produced in each of the two speaking styles was also calculated for comparison.

3. Results

All the acoustic measures were presented for each gender for the two groups of speakers in each of the two speaking styles. For vowel duration, F0, and percentage change of pitch range, data presented were pooled over the four point vowels. For these three measures, the data were submitted to a mixed-design ANOVA with Speaker Group (SG), Cantonese (Cant) and English (Eng), as a between-subjects factor, and Speaking Style (SS), Clear (Clr) and Conversational (Con), as a within-subjects factor. For the two-dimensional formant space, comparison was made to examine the vowel space expansion for each individual subject group in each of the two speaking styles.

3.1. Vowel duration

Mean vowel duration (ms) in speech produced clearly and conversationally by the female speakers and the male speakers are shown in Figure 1 and Figure 2, respectively.

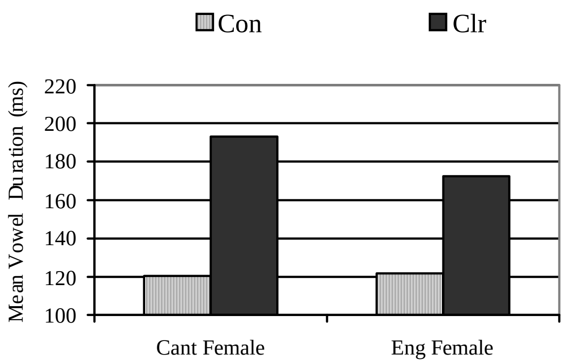

Figure 1: Mean vowel duration (ms) spoken clearly and conversationally by the female speakers.

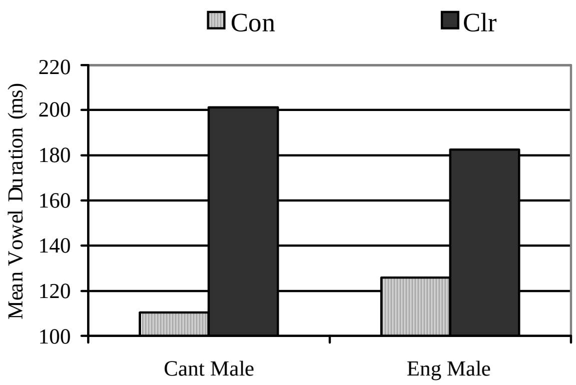

Figure 2: Mean vowel duration (ms) spoken clearly and conversationally by the male speakers.

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after abstract only review

For both female and male speakers, the analyses yielded significant effects of SS, *Fs* (1, 8) = 31.22 and 15.00, *ps* < 0.001, indicating that the vowels were significantly longer in clear speech than in conversational speech. However, within each gender, neither the effect of SG nor that of the SG x SS interaction approached significance (*ps* > 0.05), indicating that the vowel durations produced by the Cantonese speakers of English and native English speakers were not significantly different in each of the two speaking styles, and that both groups of speakers (Cant and Eng) performed similarly in each style.

3.2. Fundamental frequency

Figure 3 and 4 show the mean values of fundamental frequency (Hz) in vowels spoken clearly and conversationally by the female and the male speakers, respectively.

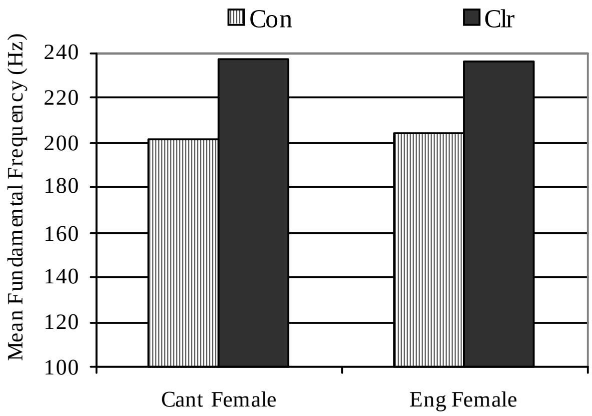

Figure 3: Mean fundamental frequency values (Hz) spoken clearly and conversationally by the female speakers.

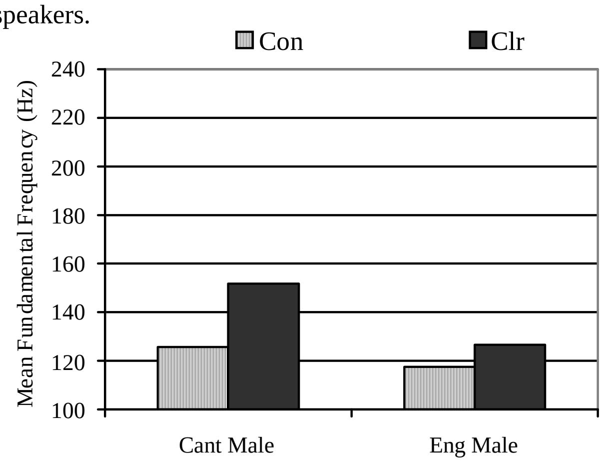

Figure 4: Mean fundamental frequency values (Hz) spoken clearly and conversationally by the male speakers.

For mean values of fundamental frequency, only the main effects of SS were found to be significant for female speakers, *F* (1, 8) = 12.41, *p* < 0.01, and for male speakers, *F* (1, 8) = 10.41, *p* < 0.05. In contrast, neither the effects of SG nor the interaction of SG x SS were found to be significant within each gender (*ps* > 0.05). The results illustrated that all the speakers exhibited mean Fo values in vowels that were higher in a clear-speaking style than in a conversational manner. In addition, the analysis suggested that differences in mean Fo values in vowels produced by the English-speaking Cantonese speakers and the native English speakers in each speaking style were non-significant, and that they increased their mean fundamental frequencies to a similar extent in clear speech

3.3. Percentage change of pitch range

The mean pitch range (% change) in vowels spoken clearly and conversationally by the female speakers and the male speakers are shown in Figure 5 and Figure 6, respectively.

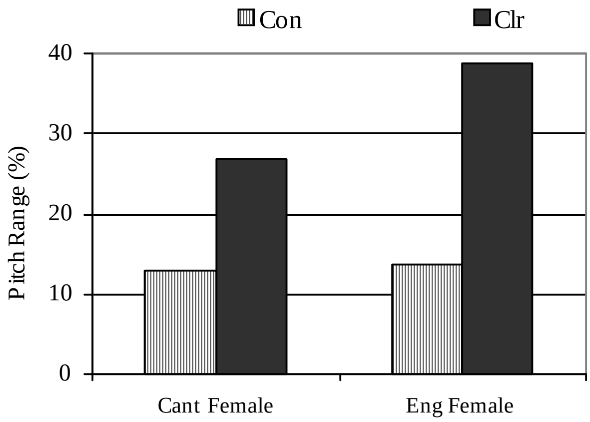

Figure 5: Mean percentage change of pitch range spoken clearly and conversationally by the female speakers.

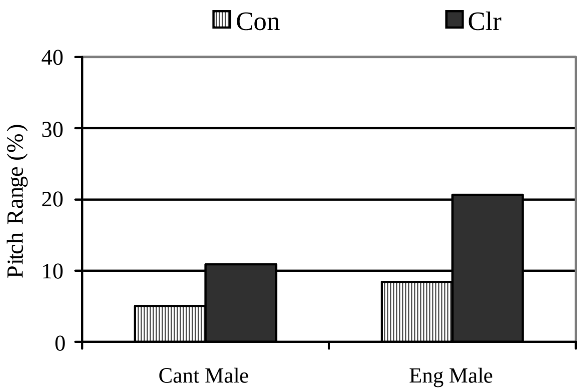

Figure 6: Mean percentage change of pitch range spoken clearly and conversationally by the male speakers.

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after abstract only review

As were the cases in vowel duration and mean values of F0, only the main effects of SS on the percentage change of pitch range reached significance, F (1, 8) = 12.43, $p < 0.01$, for the female speakers, and F (1, 8) = 31.22, $p < 0.001$, for the male speakers, indicating that all the speakers expanded their F0 range in clear speech to a greater extent than in conversational speech. As in the other two acoustic measures, no significant differences in the pitch range between the native and non-native English speakers in each gender were found, nor was there any significant interaction of SG and SS.

3.4. Vowel space area

Figures 7, 8, 9, and 10 show the F1-F2 vowel space produced in clear speech and in conversational speech by the female Cantonese speakers, female English speakers, male Cantonese speakers, and male English speakers, respectively.

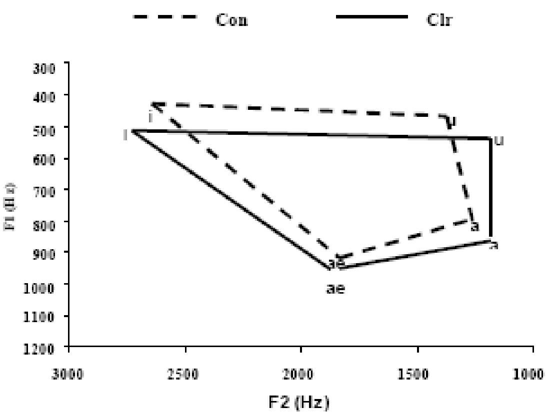

Figure 7: F1-F2 vowel space of the Cantonese female speakers in clear and conversational speech.

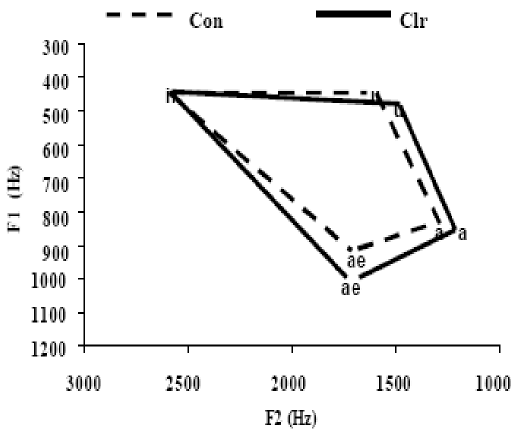

Figure 8: F1-F2 vowel space of the English female speakers in clear and conversational speech.

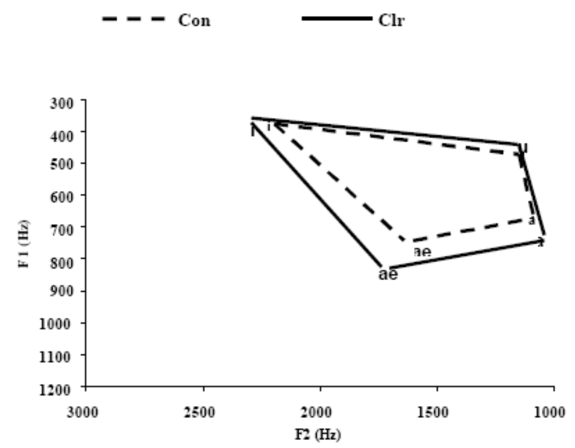

Figure 9: F1-F2 vowel space of the Cantonese male speakers in clear and conversational speech.

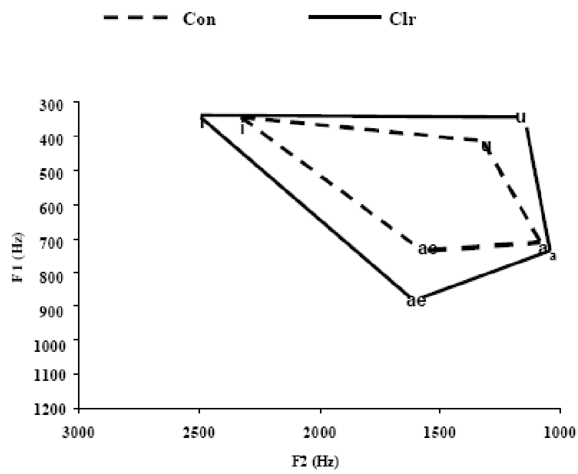

Figure 10: F1-F2 vowel space of the English male speakers in clear and conversational speech.

It has been found the vowel space areas for all the speaker groups were expanded in the clear-speaking mode. In general, the formant space areas for the female speakers were increased to a lesser extent than those for the male speakers. The vowel space expansion for the female Cantonese speakers of English was 15%, while that of the female English speakers was 20%. For the male speakers, the vowel space expansion for the Cantonese speakers was 42%, and that for the native English speakers was 69%, the greatest among the four subject groups.

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9

Accepted after abstract only review

4. Discussion and conclusions

As in previous studies that have examined the acoustic properties of clear versus conversational speech, both native and non-native speakers of English, upon given the simple instruction to “speak as clearly as possible”, exhibited the clear speech characteristics. More importantly, non-native speakers of English employed the same clear speech production strategies used by native English speakers.

All speakers produced longer vowel durations, higher mean values of fundamental frequency, greater pitch range, and larger vowel space areas in clear speech compared to conversational speech. Also, no significant differences were found in the first three acoustic measures (i.e., vowel duration, mean F0, and percentage change of pitch range) between the non-native English speakers and native English speakers, as far as the four point vowels being examined in the present study were concerned. It has been well documented that all these acoustic characteristics contribute to the intelligibility benefit of clear speech over normal, conversational speech. The findings of the present study could have important implications for improving speech intelligibility of ESL speakers, as they can be instructed to enunciate carefully. No less important is to remind health care professionals (e.g., physicians, nurses, audiologists, etc.) whose first language is not English to speak as clearly as possible when communicating with native English speakers, especially with older adults with or without hearing impairment.

5. References

Bradlow, A., & Bent, T. (2002). The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America, 112*(1), 272-284.

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

Ferguson, S. (2004). Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. *Journal of Acoustical Society of America, 116*(4), 2365-2373.

Ferguson, S., & 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*(1), 259-271.

Iverson, P. & Bradlow, A. (2002). The recognition of clear speech by adult cochlear implant users. In S. Hawkins & N. Nguyen (Eds.), *Temporal Integration in the Perception of Speech* (p. 78). Cambridge, UK: Centre for Research in the Arts, Social Sciences, and Humanities.

Krause, J. & Braida, L. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *Journal of the Acoustical Society of America*, *115*(1), 362-378.

Loizou, P. (1999). COLEA: A MATLAB software tool for speech analysis. Retrieved on October 1st, 2003 from http://www.utdallas.edu/~loizou/speech/colea.htm.

Payton, K., Uchanshki, R., & Braida, L. (1994). Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *Journal of the Acoustical Society of America, 95*(3), 1581-1592.

Picheny, M., Durlach, N., & Braida, L. (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., Durlach, N., & Braida, L. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, *29*, 434-446.

Schum, D. (1996). Intelligibility of clear and conversation speech of young and elderly talkers. *Journal of the American Academy of Audiology, 7*(3), 212-218.

Smiljanic, R., & Bradlow, A. (2005). Production and perception of clear speech in Croatian and English. *Journal of Acoustical Society of America*, *118*(3), 1677-1688.

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after abstract only review