Effects of Vowel Acoustics on the Intelligibility and Comprehensibility of China English

Qian Wang
The Chinese University of Hong Kong (Shenzhen)
wangqian@cuhk.edu.cn

ABSTRACT

The present study examines the impact of vowel acoustic properties on the perceived intelligibility and comprehensibility of China English (CE), an emergent, fast-growing, yet underdescribed variety of English spoken primarily by L1 Mandarin speakers. The acoustic analysis reveals that certain vowel distinctions present in Inner-Circle English dialects are less prominent in CE, while vowel length contrasts remain preserved within the CE vowel system. Additionally, there is a tendency to monophthongize the diphthongs in CE. Multiple regression analysis was conducted to examine the effects of vowel quality contrast, vowel length contrast, degree of vowel centralization, and degree of monophthongization on listeners' vowel recognition performance and comprehensibility rating. The results suggest that the length contrast between /ɑː/ and /ʌ/ and the degree of monophthongization of /eə/ and /oʊ/ significantly impact the perceived intelligibility and comprehensibility of CE.

Keywords: vowel acoustics, intelligibility, perception, comprehensibility, China English

1. INTRODUCTION

In today's globalized world, where English is increasingly used as a lingua franca, many new English varieties are emerging in different parts of the world. China English (CE), as one of the fastest-growing varieties of English with the largest number of speakers, has gained a great deal of attention in recent years [7, 9, 12]. Although the intelligibility of native Englishes has been studied extensively over the past few decades, remarkably little research has been conducted on the intelligibility of CE. Nevertheless, research has been conducted on the intelligibility of Chinese learners of English, revealing that CE exhibits distinct phonological features that can influence listeners' perceptual judgments [1, 8, 14, 17].

One of the most contentious topics in English as a lingua franca (ELF) research is the extent to which a new English variety is intelligible to international listeners, as it plays a critical role in determining the legitimacy of the variety. Levis [13] outlined the distinction between broad and narrow views of intelligibility. The former defines intelligibility as the accuracy of listeners' orthographic transcriptions of L2 speech, while the latter refers to listeners' impressionistic perception of how easily they understand L2 speech. Munro and Derwing [15] refer to the broad sense of intelligibility as comprehensibility, and they pointed out that these two constructs did not measure the same thing because it is common that listeners can transcribe the speech perfectly and yet perceive it to be difficult/effortful to understand. In this study, both narrow intelligibility and broad intelligibility (i.e., comprehensibility) were measured via a forced-choice word identification task and a scalar rating task, respectively.

Previous studies examining the factors that make non-native speech unintelligible suggest that pronunciation has a dominant role in the loss of intelligibility of non-native speech. Derwing and Rossiter [5] claim that prosody severely impairs intelligibility. On the other hand, other studies have emphasized the importance of segmental features, such as vowel and consonant sounds, in causing unintelligibility [16]. Jenkins [10] argued that pronunciation issues account for the biggest source of loss of comprehensibility and intelligibility, and these most commonly occur at the segmental level. In Jenkins' lingua franca core, vowel length contrast (especially before voiced/unvoiced consonants) and vowel quality of /ɜː/ were core pronunciation features that hindered intelligibility [10].

Previous studies have reported that certain vowel characteristics in CE have the potential to impact both the intelligibility and comprehensibility of CE. Hung [9], in his seminal work on the phonology of CE, found that CE has a simplified system of 6-7 vowels with a general lack of long/short or tense/lax contrast and diphthong simplification. Previous research has shown that variability in intelligibility can be related to differences in vowel space area [2,3], i.e., speakers who have larger vowel spaces tend to be more intelligible.

Given the above-mentioned, the study aims to investigate the impact of CE speakers' vowel production on the intelligibility and comprehensibility of CE. In particular, this study analyzes vowel quality contrast, vowel length

contrast, degree of vowel centralization, and degree of monophthongization of CE diphthongs and their effect on CE's intelligibility and comprehensibility.

2. METHOD

The study had two interrelated phases: production (identifying CE's acoustic features) and perception (examining listeners' recognition and evaluation of the speaker's comprehensibility) to establish the link between CE's vowel realizations and listener perception.

2.1. Participants

CE is defined as the English variety spoken by well-educated and highly proficient Chinese speakers of English. According to this working definition, six female CE speakers, aged 19-24 years (mean age of 21.5), were recruited from a high-caliber university where English is the primary language of instruction. They had been learning English for 14-16 years (mean of 14.67). Female speakers were chosen to maintain homogeneity and avoid confounding acoustic patterns caused by gender.

A total of 30 adult listeners with normal speech and hearing participated in the perception experiment. They were included into five groups based on their national backgrounds and Kachru's three-circle model of World Englishes [11]: six from Inner Circle (US), six from Outer Circle (Germany, Pakistan, India, Nigeria, South Africa), six from Expanding Circle (Vietnam, South Korea, Thailand, Poland, Myanmar), six Mandarin-speaking listeners from mainland China, and six from Indonesia.

2.2. Speech task

A word list with 19 permissible vowel phonemes in CE (five long vowels, six short vowels, and eight diphthongs) in the /hVd/ frame was used to test the intelligibility of CE vowels. Uniform word frames were used to eliminate other factors that could affect word recognition. Speakers read the words with the same vowel as the test words and then read the test words in the sentence frame "I say ___ again." Each word was read twice at a normal speed.

2.3. Listening task

Vowel intelligibility was gauged through a forced-choice word identification task. In this task, participants were asked to listen and identify each word they heard by selecting the correct word from a list of 19 words containing the target vowels and beginning with "h" and ending with "d." Before listening to the test items, each participant listened to three practice items to familiarize themselves with the task and computer interface. Each stimulus was played a maximum of two times, and all the stimuli were presented in random order for each of the listeners. Intelligibility was measured by the ratio of correctly identified words to the total number of words in the study, which is also called the "percent of correct identification" (PCI). After completing the word identification task for each speaker, listeners were asked to report their level of difficulty understanding the speaker's speech on a scalar of 1 (extremely hard to understand) to 9 (extremely easy to understand), which is how comprehensibility was measured in this study.

2.4. Acoustic measurements

The study measured several acoustic properties of vowels, including formant frequency (F1 and F2), duration contrast, quality contrast (as indicated by the Euclidean distance between four vowel pairs), and vowel space area. For diphthongs, the degree of formant movement was measured by calculating the rate of change of formant frequency values (ROC). F1 and F2 values were measured at the temporal midpoint of the monophthongs using Praat and were converted into a Bark scale [20] after obtaining each speaker's average F1 and F2 values.

(1) $Z = 13 \arctan(0.00076\,F) + 3.5 \arctan(F/7500)^2$

In this formula, F is the frequency in Hertz, and Z refers to the frequency in Bark. The vowel chart of CE plotted using the Bark values of F1 and F2 better reflects the relative perceptual distance between any two vowels. The comparison between vowels in CE and other varieties of English can be easily made in the vowel chart by using this scale.

To measure the quality contrasts between tense/lax vowels, the average Euclidean Distance (ED) was calculated between two vowels within each tense/lax vowel pair, based on their F1 and F2 values plotted in a two-dimensional vowel space. For example, to determine the ED between /i/ and /ɪ/, the two vowels are plotted as points in the vowel chart, with /i/ represented by (F1i, F2i) and /ɪ/ represented by (F1ɪ, F2ɪ). The ED can be calculated using the Pythagorean theorem, expressed as:

$$(2)\ ED(i, ɪ) = \sqrt{(F1_i - F1_ɪ)^2 + (F2_i - F2_ɪ)^2}$$

After plotting all the monophthongs in a two-dimensional vowel chart, the CE vowel space area (area of the vowel polygon based on four corner vowels /i:, u:, ɑ:, æ/) was calculated for each CE

speaker and used to indicate vowel space area and dispersion [3, 6].

To determine vowel duration, the onset and offset of each vowel were manually determined from the waveform using Praat [2, 18], with reference to a spectrogram. To compare vowel durations, mean long-to-short vowel duration ratios were calculated instead of using absolute values due to expected differences in articulation rate.

To measure diphthongs, the rate of change (ROC) of formant values (F1 and F2) was calculated using the formula "y (D-C)/x (B-A) (Hz/sec)", as shown in Figure 1 (adapted from [19]). In this figure, the vertical axis represents the formant frequency of a vowel, and the horizontal axis refers to the duration of the vowel. Point A refers to the point in time where t=20% of the entire vowel duration; Point B refers to the point where t=80% of the entire vowel duration; Point C refers to the formant value at Point B; and Point D refers to the formant value at Point A. Therefore, the ROC can be calculated by "y (D-C)/x (B-A) (Hz/sec)."

2.4. Statistical analysis

In the perception study, the acoustic features of CE vowels discussed above, including vowel quality contrast (i.e., ED), vowel length contrast (i.e., DR), degree of vowel centralization (i.e., vowel space area), and degree of monophthongization (i.e., ROC), were submitted to statistical analysis to explore how these acoustic features of vowels affect listeners' perception of CE. Stepwise multiple regression analysis was performed to further explore the degree to which the acoustic properties of vowels predicted the intelligibility and comprehensibility of CE. Stepwise regression enables the building of good models when many predictors are available, particularly if correlated. The alpha level for significance for all the tests was set as $p < 0.05$.

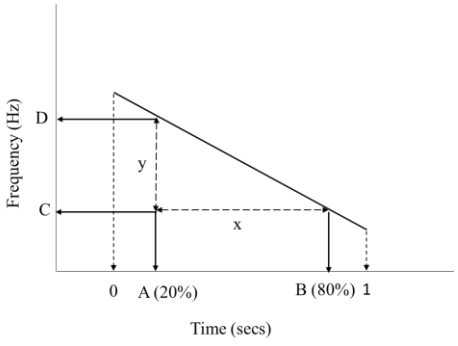

Figure 1: Schematic representation of how the ROC (rate of change) was obtained

3. RESULT

3.1. Production phase

The acoustic analyses of CE vowels show that there are some distinctive features of CE. Specifically, some vowel distinctions that exist in Inner-Circle English varieties are less evident in CE (as illustrated in Figure 2), while vowel length contrasts are still maintained. Furthermore, the study revealed a tendency to monophthongize the diphthongs in CE.

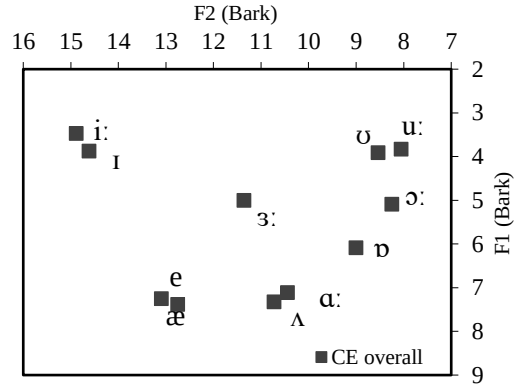

Figure 2: Formant plot for average CE monophthongs

The acoustic analyses of CE vowels revealed that vowel length contrasts were maintained in CE, with all long vowels being longer than their short counterparts (as shown in Figure 3). Paired T-tests confirmed that the durational differences between long and short values were statistically significant, except for the pair /e/-/æ/. Specifically, /iː/ was significantly longer than /ɪ/, $t(5) = 3.10$, $p < .05$; /ɑː/ was significantly longer than /ʌ/, $t(5) = 2.72$, $p < .05$; /ɔː/ was significantly longer than /ɒ/, $t(5) = 3.36$, $p < .05$; and /uː/ was significantly longer than /ʊ/, $t(5) = 4.71$, $p < .05$. However, there was no significant difference in vowel length between /e/ and /æ/ (Mean = 0.17, SD = 0.04), $t(5) = 1.85$, $p > .05$). On average, long vowels were around 1.34 times the length of short vowels.

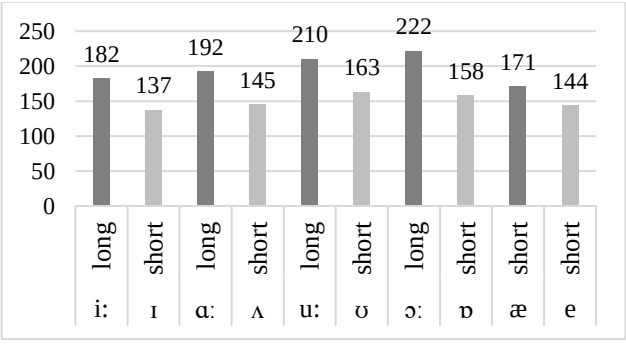

Figure 3: Duration (ms) of long and short vowels in CE

The data presented in Figure 4 indicates that the six CE speakers had varying degrees of vowel expansion, as reflected by the size of their vowel space area. The vowel space area for the CE speakers in this study ranged from 15.18 to 40.64, with S4 having the largest area (40.64) and S5 having the smallest (15.18). However, there were no significant differences in vowel space area between S1 (29.85), S2 (20.18), and S6 (24.21).

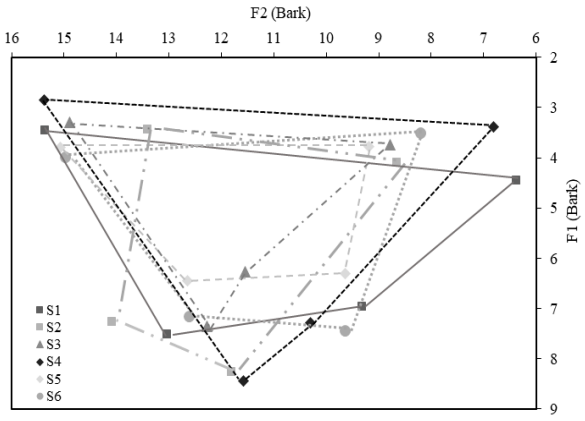

Figure 4: Vowel polygons of six CE speakers

3.2. Perception phase

The results of the statistical analyses indicate that two acoustic variables have a significant impact on the intelligibility scores. These variables are the duration ratio between /ɑː/ and /ʌ/ (Beta = .37, $p < .05$) and the degree of monophthongization of /oʊ/ as in "goat" (Beta = .16, $p < .05$), as presented in Table 1. Specifically, the study found that the intelligibility scores were positively correlated with a higher duration ratio between /ɑː/ and /ʌ/ and a larger ROC of /oʊ/. This suggests that the length contrast between /ɑː/ and /ʌ/ and the degree of monophthongization of /oʊ/ are important factors in speech intelligibility.

Predictor variable	Beta	Partial correlation	t	Sig.
Duration ratio of /ɑː/ and /ʌ/	.37	.38	5.40	.000
ROC (F2) of /əʊ/	.16	.17	2.31	.022

Note. Final model R^2 = .17, F (2, 179) = 18.59, $p <$ .01, Adjusted R^2= .16

Table 1: Result of multiple regression analysis using vowels variables as predictors of intelligibility

The results of the regression analysis indicate that only two variables had a significant impact on comprehensibility: the duration ratio between /ɑː/ and /ʌ/ (Beta = .29, $p < .05$) and the ROC of /eə/ (Beta = -.19, $p < .05$). These findings suggest that the length contrast between /ɑː/ and /ʌ/ and the degree of monophthongization of /eə/ are significant factors affecting comprehensibility ratings.

Predictor variable	Beta	Partial correlation	t	Sig.
Duration ratio of /ɑː/ and /ʌ/	.29	.29	3.96	.000
ROC (F1) of /eə/	-.19	-.18	-2.50	.013

Note. Final model R^2 =.09, F (2, 179) = 8.97, p <. 01, Adjusted R^2= .08

Table 2: Result of multiple regression analysis using vowel variables as predictors of comprehensibility

4. DISCUSSION AND CONCLUSION

The study has identified certain acoustic qualities of vowels that have a significant impact on listeners' vowel recognition performance and comprehensibility rating. These qualities include the length contrast between /ɑː/ and /ʌ/ and the degree of monophthongization of /eə/ as in "hair" and /oʊ/ as in "goat." These findings suggest that vowel length contrast and the degree of monophthongization of specific diphthongs are important factors in enhancing the intelligibility and comprehensibility of non-native speakers of English.

The study found no statistically significant impact of vowel space area and quality contrast on the intelligibility or comprehensibility ratings of CE speakers. This does not align with Bent et al. [2] and Bradlow et al. [3], which showed higher overall intelligibility for speakers with greater vowel areas. This discrepancy may be due to the high proficiency level and homogeneity of the participants. However, a closer examination of individual speakers revealed a potential positive correlation between vowel space area and word recognition ease, as S4 with the largest space area received higher comprehensibility ratings than S5 with the smallest space area.

It should be noted that these preliminary findings are limited to the word identification task and further research is needed to explore the influence of acoustic variables on the intelligibility and comprehensibility of CE speakers in natural conversation.

5. REFERENCES

[1] Bent, T., Bradlow, A. R. 2003. The interlanguage speech intelligibility benefit. *J. Acoust. Soc. Am. 114,* 1600–1610.

[2] Bent, T., Bradlow, A. R., Smith, B. L. 2008. Production and Perception of Temporal Patterns in Native and Non-Native Speech. *Phonetica 65*(3), 131–147.
[3] Bradlow, A. R., Torretta, G. M., Pisoni, D. B. 1996. Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication 20*, 255–272.
[4] Brato, T. 2018. Basic Vowel Formant Analysis for NORM (Version: 3.0). Retrieved from https://www.uni-regensburg.de/language-literature-culture/english-linguistics/staff/brato/praat-scripts/index.html.
[5] Derwing, T. M., Rossiter, M. J. 2003. The effects of pronunciation instruction on the accuracy, fluency and complexity of L2 accented speech. *Applied Language Learning 13*, 1-18.
[6] Fox, R. A., Jacewicz, E. 2008. Analysis of total vowel space areas in three regional dialects of American English. *Proceedings of Acoustics 08* (pp. 495–500). Paris, France.
[7] Hansen Edwards, J. G. 2017. China English: Attitudes, legitimacy and the native speaker construct: Is China English becoming accepted as a legitimate variety of English? *English Today 1*(2), 1–8.
[8] Hardman, J. B. 2010. *The intelligibility of Chinese-accented English to international and American students at a US university* (Unpublished doctoral thesis). The Ohio State University.
[9] Hung, T. N. 2002. *Phonological features of Chinese varieties of English.* Paper presented at the 7th English in Southeast Asia Conference, Hong Kong Baptist University, December 6–8.
[10] Jenkins, J. 2000. *The phonology of English as an international language.* Oxford: Oxford University Press.
[11] Kachru, B. B. 1985. Standards, codification and sociolinguistic realism: the English language in the outer circle. In: R. Quirk and H.G. Widdowson (Eds), *English in the world: Teaching and learning the language and literatures* (pp. 11–30). Cambridge: Cambridge University Press.
[12] Kirkpatrick, A. 2007. *World Englishes: implications for international communication and English language teaching.* Cambridge, England: Cambridge University Press.
[13] Levis, J. 2006. Pronunciation and the assessment of spoken language. In R. Hughes (Ed.), *Spoken English, TESOL and applied linguistics: Challenges for theory and practice* (pp. 245–270). New York: Palgrave Macmillan.
[14] Munro, M.J., Derwing, T.M. 1995. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning 45*(1), 73–97.
[15] Munro, M. J., Derwing, T. M. 1999. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. Article reprinted in *Language Learning 49*(1), 285-310.
[16] Munro, M. J., Derwing, T. M. 2006. The functional load principle in ESL pronunciation instruction: An exploratory study. *System 34*(4), 520–531
[17] Rogers, C. L. 1997. *Intelligibility of Chinese-accented English* (Unpublished doctoral dissertation). Indiana University, Bloomington.
[18] Smith, B. L., Hillenbrand, J., Ingrisano, D. 1986. A comparison of temporal measures of speech using spectrograms and digital oscillograms. *Journal of Speech & Hearing Research 29*(2), 270–274.
[19] Tsukada, K. 2008. An acoustic comparison of English monophthongs and diphthongs produced by Australian and Thai speakers. *English World-Wide 29*(2), 194–211.
[20] Zwicker, E., Terhardt, E. 1980. Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. *J. Acoust. Soc. Am. 29*(5), 1523–1525.