**Title**

A corpus-based analysis of vowel production of L1-Chinese learners of English

**Short title**: Vowel production of L1-Chinese learners of English

**Authors**

Martin Schweinberger\*

The University of Queensland

Address: School of Languages and Cultures, Gordon Greenwood Building, Union Road,

The University of Queensland, St Lucia, QLD 4072, Australia

Phone: +61 7 3365-6374

Email: [m.schweinberger@uq.edu.au](mailto:m.schweinberger@uq.edu.au)

ORCID: 0000-0003-1923-9153

Ruihua Yin

The University of Queensland

Address: School of Languages and Cultures, Gordon Greenwood Building, Union Road,

The University of Queensland, St Lucia, QLD 4072, Australia

Email: [ruihua.yin@uq.edu.au](mailto:ruihua.yin@uq.edu.au)

ORCID:

\*corresponding author

**Abstract**

This study combines acoustic phonetics with computational and applied corpus linguistics to analyse and compare the production of the monophthongal vowels of 148 L1-Chinese learners (CHN) and 107 L1-speakers of English (ENS) based on The *International Corpus Network of Asian Learners of English* (ICNALE). The study aims to ascertain if CHN show systematic differences in vowel duration due to the absence of phonemic duration contrasts in Chinese substrates. The study uses mixed-effects linear regression modelling to determine if CHN exaggerate durational contrasts. The analysis finds that CHN extend the duration of all vowels and exaggerate the difference between long and short vowels to compensate for the lack of qualitative differences between short and long vowel pairs. This study represents the first corpus-based acoustic analysis of CHN vowels in spontaneous speech.

**Keywords**: acoustic phonetics, vowel production, learner corpus research, Chinese learners of English

1. **Introduction**

* Chinese English learners: important group
* large number of learners (Deterding 2017, 17)
* economically important (cite numbers for Australia)

While pronunciation poses a challenge for language learners (Gilakjani and Ahmadi, 2011), it is also the most immediate and direct display of linguistic proficiency. Listeners automatically and subconsciously categorize and infer judgments about speakers based on pronunciation (Flege, 1995). In addition, pronunciation is crucial for intelligibility and is affecting real-life opportunities (jobs, partner choice, etc.).

An underlying cause for the difficulties that learners face is that languages are not independent but interact in the minds of multilingual speakers (Flege, 1995) which means that the L2 sound system is affected by the L1 system (and vice versa). From the perspective of CHN, English vowels are particularly challenging due to the absence of durational contrasts in Chinese substrates (Deterding 2017, 20; (Duanmu 2007: 41; Zhang 2009, 142), typological distance, and differences between the monophthongal vowel inventories.

left panel: vowel chart showing monophongal vowels of standard Chinese; center panel: vowel chart of Southern Californian American English; right panel: tongue position corresponding to selected front vowels.

Differences in how vowels are differentiated (Chinese: formant and tone differences versus English: formant and duration differences)

left pane: durations of high front vowels produced by ENS; right pane: durations of high front vowels produced by CHN.

Deterding (2017: 27): “Jenkins (2000: 145) has suggested that vowel length is important for maintaining intelligibility in international English, both in distinguishing long and short vowels and in signalling the voicing distinction at the end of a word. However, extract 30 is the only token I can find in the Guangxi data where vowel length might have contributed to a misunderstanding. Deterding (2013: 70) similarly concludes that vowel length is not important for maintaining intelligibility in English as a Lingua Franca interactions.”  
Deterding (2017: 30): “It is hard to draw firm conclusions based on the small number of tokens analysed here,”

Target-likeness, processing, and usage-based approach

Formants

Formants are concentration of acoustic energy at a certain frequency (Ladefoged and Johnson, 2014) with the first formant (F1) and the second formants (F2) of a vowel sound inversely corresponding to the tongue height and tongue fronting during vowel production. Regarding the production of English vowels produced by CHN, it has been shown that CHN (XXX). Furthermore, it has been reported that CHN are very sensitive to vowel duration (XXX)and exaggerate duration to compensate for the relative insensitivity to durational contrasts.

* Little empirical research on their speech production in spontaneous speech
* Existing research based on few speakers

Previous research on vowel production by CHN is predominantly based on read-aloud word lists or selected scripted sentences in highly controlled laboratory conditions. Hence, characteristics of the vowel production of learners in naturalistic speech environments remain largely unknown. Furthermore, previous research has relied on small samples of subjects with studies using between 8 and 15 subjects. As such, the findings provided by previous research may not warrant generalisation to larger speech communities or to conversational language production in natural settings.

The present study addresses these issues and aims to provide a more detailed understanding of the following research questions:

RQ1: Do CHN exaggerate the lengths of vowels in English?

RQ2: Do CHN to exhibit a lack of spectral differentiation (vowel mergers)?

1. **Previous Research**

Chinese learners of English may struggle with several vowel sounds that are different between the two languages. Some of the English vowel sounds that are particularly challenging for Chinese learners include:

1. /ɪ/ as in "sit" - This sound does not exist in Chinese, and Chinese learners may substitute it with the vowel sound /i/ as in "see", which can lead to mispronunciations like "seat" instead of "sit" (see Zhang 2009)
2. /æ/ as in "cat" - This sound is also absent from Chinese leading to difficulties in producing this vowel among Chinese learners of English (Zhang & Lu 2012). Chinese learners may substitute it with a similar but incorrect sound, such as /e/ as in "met" (Wang 2008), leading to mispronunciations like "met" instead of "mat" (see Jiang & Zhang 2019).
3. /ʌ/ as in "cut" - This sound is not present in standard Mandarin Chinese, and Chinese learners may struggle to produce it accurately, sometimes substituting it with a similar but incorrect sound such as /ɑ/ as in "father", leading to mispronunciations like "cot" instead of "cut".
4. /ɔː/ as in "law" - This sound is absent from many Chinese dialects, and Chinese learners may have difficulty distinguishing it from other English vowels, sometimes producing it as /ɑ/ as in "father" or /oʊ/ as in "low", leading to mispronunciations like "low" instead of "law".
5. /ʊ/ as in "book" - This sound is also not present in Chinese, and Chinese learners may substitute it with the vowel sound /u/ as in "moon", leading to mispronunciations like "buck" instead of "book".

It's important to note, however, that not all Chinese learners of English struggle with the same vowels, and there can be individual variation depending on factors such as age, language proficiency, and amount of exposure to English.

* Research Gap: Important learner group but little systematic empirical research on vowel production
* Linguistic relevance: no phonemic duration contrast in monophthongal vowels
* How do learners deal with duration in English given the absence of durational contrasts in their Chinese L1 substrates?
* Existing research
  + Few speakers (generalizability?)
  + Lab settings (generalizability?)
  + Not practice oriented (no applications)

Deterding (2009, 193): “An alternative way to describe these tokens is to say that the vowel in in is very close, so the distinction between /i˜/ and /I/ is not always maintained, especially before a final nasal.”

Su-Hyun and Liu (2013) “These results suggest that when producing vowels /i, I/ in the /hVd/ context, EN speakers lowered their tongue position, while CN and KN speakers did not change their tongue height.”

Li and Sewell (2012:82): “The possibility of certain vowel contrasts (such as that between /e/ and /æ/) being absent from China English is mentioned by Ma (2007), who relates them to the different phonemic and allophonic distributions in Mandarin and English.”

Ao and Low (2012, 31): “/ʌ/ pronounced as /ɑ/ As the most salient vocalic feature of pronunciation of YE speakers, this most frequently occurs in words such as duck /dɑk/, up /ɑp/, us /ɑs/, come /kɑm/ and cousins /'kɑzǝnz/. This feature, however, needs further validation through acoustic measurement in the future study due to the fact that it is no easy task to differentiate aurally between these two sounds. This feature is found in the pronunciation of nine of ten speakers and may be considered as a wide-spread feature of YE.”

Ao and Low (2012, 32): “The features of these three vowels are hard to categorize because each of these vowels is pronounced at least in two different ways. The long vowel /iː/ is pronounced /e/ in fields /feldz/ (by two speakers) but /eɪ/ in feast /feɪst/ (by three speakers). The pronunciation of vowel /ɪ/, probably due to its complex nature, varies from word to word or even for the same word, e.g. it /jiːt/, village/villagers /'velɪʤ, 'veɪlɪʤ/ and fist /fiːst, feɪst, fɜːst/. The vowel /e/, on the other hand, is realized in three different ways, namely /ɜː/, /iː/ and /eɪ/, as in the words however /hɑʊ'ɜːvǝ/, threaten /'θriːtǝn, 'sriːtǝn/, pleasure /'pleɪjǝ/ and successful /sǝk'seɪsfɔː/.”

Ao and Low (2012, 29): “but its [CE] pronunciation features have not received equal attention. In fact, to date, the only detailed and systematic empirical studies in this area are Hung (2002) and Deterding (2006b). Hung's (2002) study analyzed the speech of 93 first-year undergraduate non-English-major students at three universities in Beijing who are from 10 Chinese dialect groups, namely Beijing, Tianjin, Northeast, Northwest, Yue (Cantonese), Wu, Min (Fukien), Kejia (Hakka), Gan (Jiangxi) and Xiang (Hunan). All students were recorded reading 100 words in citation forms (e.g. ‘This is a worry’; ‘This is a cord’). The phonetic features of Chinese English identified in the study were categorized under each Chinese dialect group.

Deterding (2006b) analyzed and described the pronunciation of 13 non-English-major undergraduates originally from nine Chinese provinces, namely Anhui, Henan, Hunan, Jilin, Jiangsu, Jiangxi, Liaoning, Shandong and Zhejiang. Each student was recorded reading a passage and participating in a short interview. He identifies some salient features of pronunciation which he argues ‘may become established as part of a unique variety of English that is emerging in China’ (Deterding, Reference Deterding2006b: 175). There are a few other studies of CE pronunciation features, which include Ho (Reference Ho, Lee, Ho, Meyer, Varaprasad and Young2003), who videotaped and transcribed oral presentations of 39 PRC students who were studying in Singapore, Chang (Reference Chang, Swan and Smith1987), Pride and Liu (Reference Pride and Liu1988), Jiang (Reference Jiang2002), and He and Li (Reference He and Li2009) who provide some impressionistic descriptions of phonological features of Chinese English as a whole.”

1. **Data and Methodology**
   1. **Corpus Data**

The study uses data from the *International Corpus Network of Asian Learners of English* (ICNALE)(XXX). The ICNALE is one of the largest publicly available multimodal learner corpora comprising more than 10,000 topic-controlled speeches and essays produced by college students in ten countries and regions in Asia as well as English native speakers. For this study, all data representing spoken monologues (spontaneous speech) collected between 2017 and 2019 from 148 CHN and 107 ENS were analysed (this encompasses all CHN and ENS speakers in the data that produced relevant tokens and, in the case of ENS, spoke American English).

Every speaker contributed two one-minute recording to the spoken monologues’ component of ICNALE. Speech samples were recorded on mobile devices or personal computers using in-built microphones resulting in a highly variable quality of recordings.

* 1. **Data Processing**

Data processing started with using Web-MAUS (XXX)to (force) align the audio files and transcriptions provided by ICNALE into Praat TextGrids (the forced alignment used both US and British models). All subsequent steps of the analysis were performed using R Version 4.2 (XXX) in RStudio(XXX).

The first to third formants of all vowels as well as vowel durations were extracted using rPraat (XXX), wrassp (XXX), and tidyverse (XXX). The algorithm targeted a range between three and 7 formants for each vowel resulting in five formant values for each of the first to third formants in each vowel. The optimal formant values out of these five options were determined based on the minimal Euclidean distance to standard American English vowel formants based on (XXX) and standard southern British English based on (XXX). Based on the Euclidean distance, alternative options and the less well fitting target variety were removed from the analysis so that the data set contained only one observation for each vowel (see Table 1)

Overview of the semi-processed data set.

In a next step, socio-demographic information about the speakers (speaker type, age, gender, English proficiency) was added to the data and ENS not from North America were removed. All further analyses continued with standard American English as target variety. Table 2 provides an overview of the socio-demographics of the speakers in the data and Table 3 shows the proficiency levels among the L1 Chinese learners.

Overview of information about the speakers represented in the data.

Overview of proficiency levels among CHN speakers in the data.

Next, vowels were normalized using a z-transformation after grouping the data by speaker type (ENS vs CHN) and gender. After this normalization procedure, all vowels not representing i:/, I/, u:/, and U/ were removed from the analysis.

Then, multi-syllabic words or words containing more than 9 characters were removed to better control for variability caused by the phonetic and phonological environments in which the vowels were produced. Only words were retained which had a /CV(C)/ syllable structure (e.g., *get*, *gut*, *hit*, *shit*, *due*, *we*, *see*).

To account for the low quality of audio recordings and to remove outliers and inaccuracies, kernel density estimation was applied to the z-transformed first and second formats. All vowels having density values in the lower quartile of first and second formats were removed. The final data set is summarized in Table 4.

Overview of the final data set (percentage of retained observations compared to semi-processed data in brackets).

* 1. **Statistical Analysis**

The statistical analysis made use of two procedures

Bhattacharya coefficients: to assess potential spectral mergers of i:/ and I/ as well as u:/ and U/

Mixed-effects linear regression: to assess if CHN exaggerate the length of vowels to compensate for a potential lack of spectral differentiation

Bhattacharya coefficients are suited to assess vowel mergers as this coefficient represents a measure of overlap of scatter clouds with 1 representing perfect overlap and 0 representing zero overlap.

Mixed-effects linear regression modelling was performed using the lme4 (XXX)and the sjPlot package (XXX) with a step-wise step-up model fitting procedure. The regression analysis evaluated the effect of the following variables and their two-way interactions. If models exhibited substantial multicollinearity (variance inflation factors > 5, the model was considered not trustworthy). Table 5 details the variables that were tested during the statistical modelling and provides information about their scaling as well as how they were operationalized.

Variables included in the mixed-effects regression modelling (ran. eff. = random effect, ind. var. = independent variable, cat. = categorical scaling, nom. = nominal scaling).

Proficiency could not be included into the model and used as a predictor as no proficiency information was available for ENS. Including proficiency as a predictor would have led to the automatic exclusion of all ENS data points or it would have resulted in ill-fit models due to the absence of variability in proficiency among ENS.

1. **Results**

The following reports on the findings of the analysis separated by research question the statistical analyses have addressed.

* 1. **Vowel duration**

The regression modelling arrived at a final minimal adequate model with a notable explanatory capacity accounting for XXX percent of the overall variability in the data. The data confirmed speaker type, vowel, and gender as significant main effects. More importantly, as this directly addresses the second research question and in addition to an interaction between gender and speaker type, the model reports significant interactions between speaker type and vowels. These latter interactions confirm that CHN exaggerate all vowel durations compared with the vowel durations of ENS.

Results of the mixed-effects regression modelling.

The results show that CHN extend or exaggerate all vowel durations and not just short vowels. The effect plot shown in Figure 5 furthermore shows that CHN exaggerate the duration difference of both i:/ and I/ as well as u:/ and U/. This is confirmed by Figure 6 as the differences between durations are notably higher for both vowel pairs among CHN compared with ENS.

Predicted duration values based on the final minimal adequate model by vowel and speaker type. (Gray and orange lines show the difference between short and long vowel durations within speaker groups)

Duration differences by speaker type (orange: ENS, gray: CHN) and vowel pair. The bars show the difference in duration between i:/ and I/ (left) and between u:/ and U/ (right).

The results of the statistical analysis thus confirm that CHN merge spectrally similar vowels and appear to compensate the lack of a differentiation by exaggerating both the durations of high-front and -back vowels and by exaggerating the durational differences between long and short vowels (see Figure 6).

* 1. **Vowel duration of words**

1. **Discussion**

Summary: exaggerate all vowels (short and long)

The findings presented here confirm previous research which reported the tendency to exaggerate vowel durations produced by CHN in lab settings (XXX).

Deterding (2006, 183): “It seems that while many of these speakers use a reduced vowel in the unstressed syllables of polysyllabic words, they almost always use a full vowel in monosyllabic function words. One explanation for this pattern might be that words are often learned in isolation in China, and while a schwa is likely to occur in the citation form of a content word, it will never occur in a function word in isolation. If learning English depends substantially on memorization of isolated words, the weak forms of function words may never occur.”

The findings presented here also offer unique insights in that the study extends previous research to natural settings and substantially expands the empirical basis of existing research. In addition, the data analysed here suggests a merger of high-back vowels (u:/ and U/) among ENS in spontaneous speech.

A noteworthy limitation of the present study relates to the variable quality of the recordings which can be considered not only substandard but relatively poor for at least a subsection of the minute-long recordings that make up the spoken monologue component of the ICNALE. The reason for the low quality of the audio recordings is that the audio data were recorded predominantly using in-built microphones of mobile devices. While the quality of the audio data could, at least in part, be compensated using statistical procedures (kernel density estimation) which reduced the existing noise to a certain extent, such means are ultimately limited and unfit to fully remedy data quality.

Another limitation consists in the fact that, given the variability and distributional characteristics of spontaneous speech, it is difficult to control the semantic and phonological environments of vowels, which, however, affect vowel production and thus formant values.

* Finally, the study does not assess phonetic realizations of vowel durations in the Chinese substrates.

The advantages of the present study are that it has produced insights into vowel production by CHN in spontaneous speech which is under-explored even in learner corpus research. Also, the study is among the first to study CHN vowel production in natural settings which allows to generalize findings to real-life learner speech. Despite its limitations, the fact that the poor quality of the data could be compensated using advanced methods enables to extend the methods presented here to further automated corpus-based investigation on larger and more diverse samples of learner speech.

1. **Conclusions**

The present study adds to the relatively new body of research with takes a large-scale, corpus-based approach to studying ESL vowel production in natural speech. The advantage of this approach consists in its ability to extending previous research by substantively extending the data base in terms of both the number of speakers and observations. The application of kernel density estimation to mitigate low quality audio data is promising but requires support based on comparisons against gold standard data sets as well as manually annotated data. Such comparisons will allow us to evaluate to what extent this method can compensate for poor data quality. If evaluated positively, the approach presented here can be extended to other learner varieties and multi-modal data sources. Potential follow-up studies need to evaluate the findings here in terms of perception and intelligibility to investigate the auditory and cognitive implications of the acoustic effects presented here. Finally, the present study offers insights which can easily be translated into teaching and learning resources and is thus relevant to practitioners. can be a prototype

1. **References**

Deterding, D. (2006). The pronunciation of English by speakers from China. English World-Wide, 27(2), 175-198.

Deterding, D. (2009). The formants of monophthong vowels in standard southern British English pronunciation. Journal of the International Phonetic Association, 39(1), 47-55.

Deterding, D. (2017). The pronunciation of English in Guangxi: Which features cause misunderstandings? In Z. Xu, D. He, & D. Deterding (Eds.), Researching Chinese English: The state of the art (pp. 17-32). Springer.

Duanmu, S. (2007). The phonology of Standard Chinese (2nd ed.). Oxford: Oxford University Press.

Flege, J. E. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), Speech perception and linguistic experience: Issues in cross-linguistic research (pp. 233-277). York Press.

Gilakjani, A. P., & Ahmadi, M. R. (2011). Why is pronunciation so difficult to learn? English Language Teaching, 4(3), 74-83.

Jenkins, J. (2000). The phonology of English as an international language. Oxford: Oxford University Press.

Bořil, T., & Skarnitzl, R. (2016). Tools rPraat and mPraat. In P. Sojka, A. Horák, I. Kopeček, & K. Pala (Eds.), Text, Speech, and Dialogue (pp. 367-374). Springer International Publishing.

Bombien, L., Winkelmann, R., & Scheffers, M. (2021). wrassp: An R wrapper to the ASSP Library (R package version 1.0.1).

Ishikawa, S. (2014). Design of the ICNALE spoken: A new database for multi-modal contrastive interlanguage analysis. Learner Corpus Studies in Asia and the World, 2, 63-76.

Kisler, T., Reichel, U. D., & Schiel, F. (2017). Multilingual processing of speech via web services. Computer Speech and Language, 45, 326-347.

Ladefoged, P., & Johnson, K. (2014). A course in phonetics. Cengage.

R Core Team. (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing. <https://www.R-project.org>

RStudio Team. (2022). RStudio: Integrated development environment for R. RStudio PBC. <http://www.rstudio.com>

Su-Hyun, J., & Liu, C. (2013). The vowel inherent spectral change of English vowels spoken by native and non-native speakers. The Journal of the Acoustical Society of America, 133. <https://doi.org/10.1121/1.4798620>

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1), 1-48.

Lüdecke, D. (2021). sjPlot: Data visualization for statistics in social science (R package version 2.8.10).

Tsukada, K. (2001). Native vs non-native production of English vowels in spontaneous speech: An acoustic phonetic study. In P. Dalsgaard, B. Lindberg, & H. Benner (Eds.), Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech 2001) (pp. 305-308).

Visceglia, T., Chiu-Yu, T., Kondo, M., Meng, H., & Sagisaka, Y. (2009). Phonetic aspects of content design in AESOP (Asian English Speech cOrpus Project). In Oriental COCOSDA International Conference on Speech Database and Assessments (pp. 60-65).

Wickham, H., et al. (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686.

Yang, B. (1996). A comparative study of American English and Korean vowels produced by male and female speakers. Journal of Phonetics, 24, 245-261.

Zhang, F. (2009). A study of pronunciation problems of English learners in China. Asian Social Science, 5(6), 141-146.

Li, Siqi & Sewell, Andrew. 2012. Phonological Features of China English. Asian Englishes 15 (2), 80-101, DOI: 10.1080/13488678.2012.10801331

Ao, Ran and Low, Ee Ling. 2012. Exploring pronunciation features of Yunnan English. English Today 111 28 (3). doi:10.1017/S0266078412000284

Best, C. T., Tyler, M. D., Gooding, T. N., Orlando, C. B., & Quann, C. A. (2009). Development of phonological constancy: 19-month-olds, but not 15-month-olds, identify words in a non-native regional accent. Child development, 80(3), 672-680.

Chen, M. Y., & Munro, M. J. (2008). The acoustic characteristics of English vowels produced by Chinese and Canadian English speakers. The Journal of the Acoustical Society of America, 124(1), 324-333.

Flege, J. E., & Liu, S. (2001). The effect of experience on nonnative subjects’ production and perception of English vowels. Journal of phonetics, 29(2), 217-230.

Guion, S. G., & Pederson, E. (2007). Investigating the relationship between cross-language similarity and phonetic category learning. The Journal of the Acoustical Society of America, 121(4), 2264-2277.

Wang, J., & Munro, M. J. (2004). Acoustic characteristics of English vowels produced by Chinese and English male speakers. The Journal of the Acoustical Society of America, 116(4), 2604-2615.

Chen, Y., Peng, G., & Wang, W. S. Y. (2012). Vowel production of American English by native Mandarin speakers: Acoustic analysis and perceptual data. The Journal of the Acoustical Society of America, 132(6), 4028-4039.

Chen, Y., & Wong, L. L. N. (2014). Mandarin tone and vowel effects on the production and perception of American English word stress by Chinese and American English speakers. The Journal of the Acoustical Society of America, 136(4), 1827-1838.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. Journal of the Acoustical Society of America, 107(5), 2711-2724.

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.

Wang, Y. (2018). Phonological and acoustic correlates of English vowels produced by Chinese EFL learners. Journal of Second Language Pronunciation, 4(2), 233-262.

1. **Miscellaneous**

Regarding the potential merger between i:/ and I/ among CHN and ENS speaker, the high values of the Bhattacharya coefficient confirmed the expected merger of spectrally similar high-front vowels among CHN but not among ENS:

CHN Bhattacharya coefficient (i:/, I/):

ENS Bhattacharya coefficient (i:/, I/):

left panel: overlap of high-front vowels among CHN; right panel: overlap of high-front vowels among ENS.

Regarding the potential merger between u:/ and U/ among CHN and ENS speaker, the high values of the Bhattacharya coefficient confirmed the expected merger of spectrally similar high-back vowels for CHN but also confirmed a merger of spectrally similar high-back vowels among ENS:

CHN Bhattacharya coefficient (u:/, U/):

ENS Bhattacharya coefficient (u:/, U/):

left pane: overlap of high-back vowels among CHN; right pane: overlap of high-back vowels among ENS.

Thus, the analysis confirmed the expected mergers of spectrally similar high-front and -back vowels among CHN but also a merger of high-back vowels among ENS.