A corpus-based acoustic analysis of vowel production by L1-Chinese learners and native speakers of English

Martin Schweinberger
Ruihua Yin

m.schweinberger@uq.edu.au
ruihua.yin@uq.edu.au

All materials (except data) available at https://github.com/MartinSchweinberger/ICAME44

A corpus-based acoustic analysis of vowel production by L1-Chinese learners and native speakers of English

Timeline | Table of Contents

- Background and Motivation
- Research Gaps | Research Questions
- Methodology (Data and Analysis)
- Results
- Discussion and Outlook

Background and Motivation

Pronunciation is a challenge for L2 English learners

Problem

- Pronunciation is most immediate and direct
- Everybody automatically and subconsciously categorizes and infers judgements based on pronunciation (gender, age, cultural background, nativeness, socio-economics, education, etc.)
- Pronunciation is crucial for intelligibility
- Pronunciation is affecting real-life opportunities (jobs, partner choice, etc.)

Production and perception of speech sounds are crucial for mutual intelligibility in daily converation!

Background and Motivation

Why is pronunciation a challenge for L2 English learners?

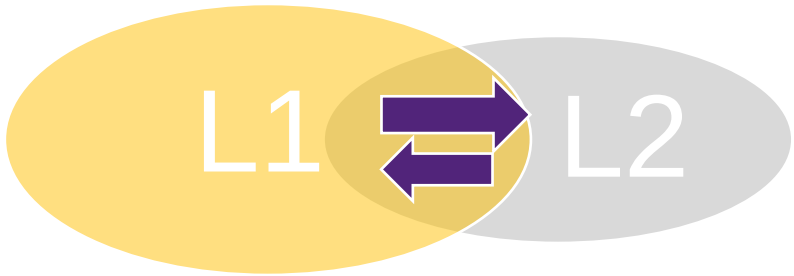
Languages interact in the minds of multilingual speakers

Language transfer (Odlin, 1989)

The influence resulting from similarities and differences between the target language and any other language that has been previously acquired.

Speech Learning Model (SLM) (Flege 1995)

- L1 and L2 sound systems exist in a shared phonetic space in the bilingual mind
- As a result, the L2 sound system is affected by the L1 system (and vice versa)



Background and Motivation

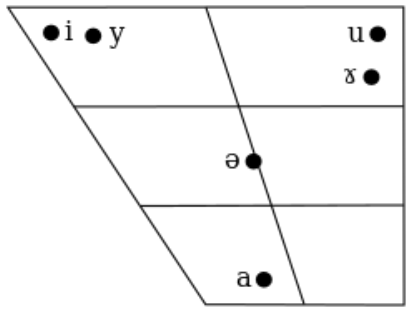
Prior research reports that Chinese learners struggle with English vowels, in particular:

- Problems differentiating between
  - /i:/ and /ɪ/ (see, e.g., Zhang & Yin, 2009)
  - /e/ and /æ/ (Wang, 2008; Zhang & Lu, 2012; Jiang & Zhang 2019)

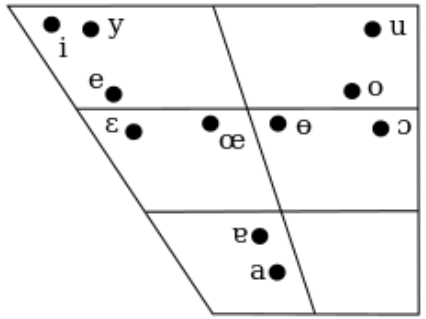
Reasons

- Differences in inventory size (CHN: Mandarin 6 vowels* vs. ENS: app.** 11 vowels) (Wang & Sun, 2015)
- Differences in how vowels are differentiated (ENS: formants + duration | CHN: forma

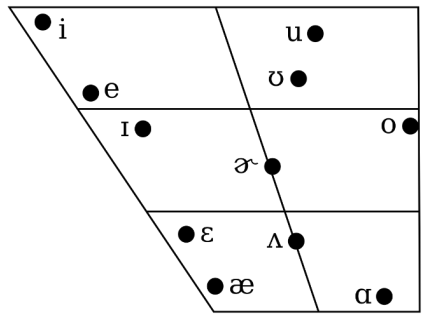
Mandarin monophthongal vowels

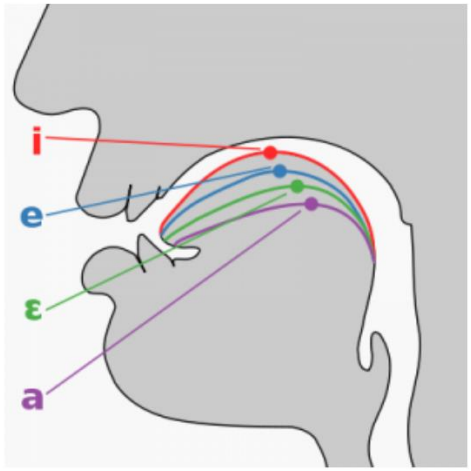


Cantonese monophthongal vowels



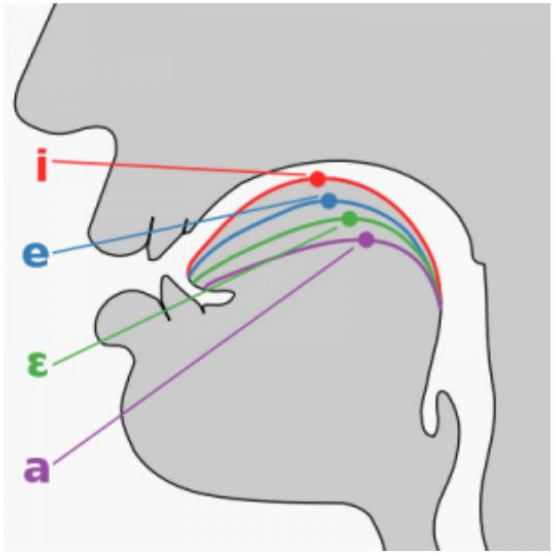
Monophthongal vowels in southern Californian Engl.





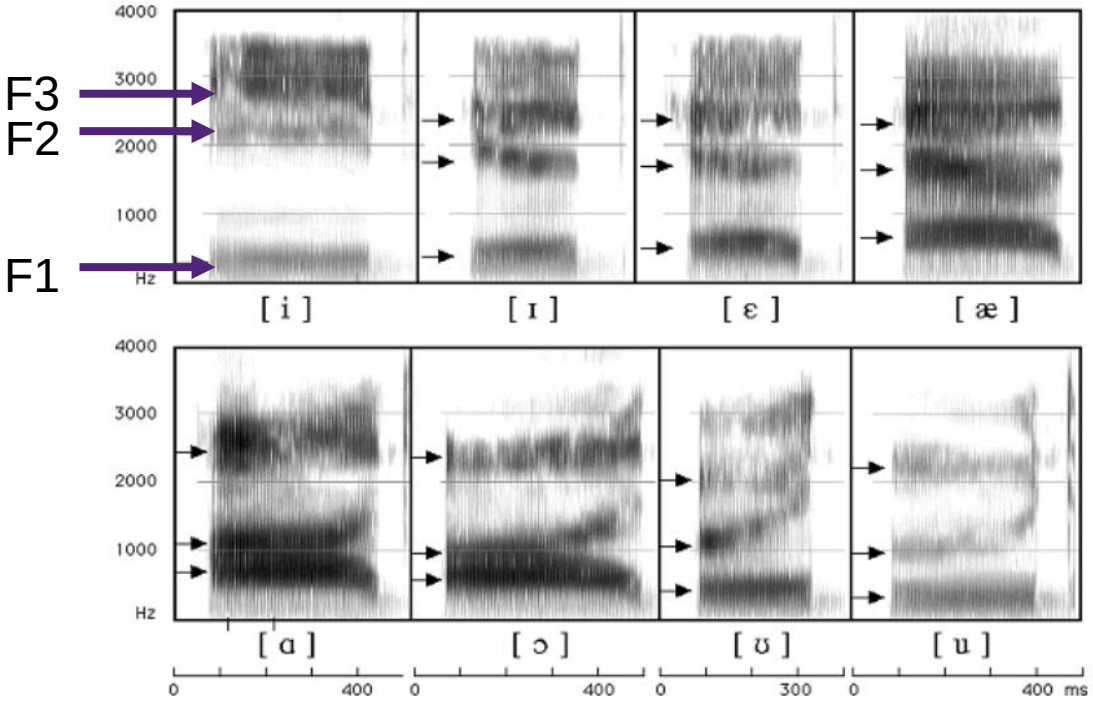
* Cantonese has 11 monophthongal vowels, i.e., vowels with a fixed tongue position (no moving tonue position like /au/ or /ou/.
** Depending on the variety of English.

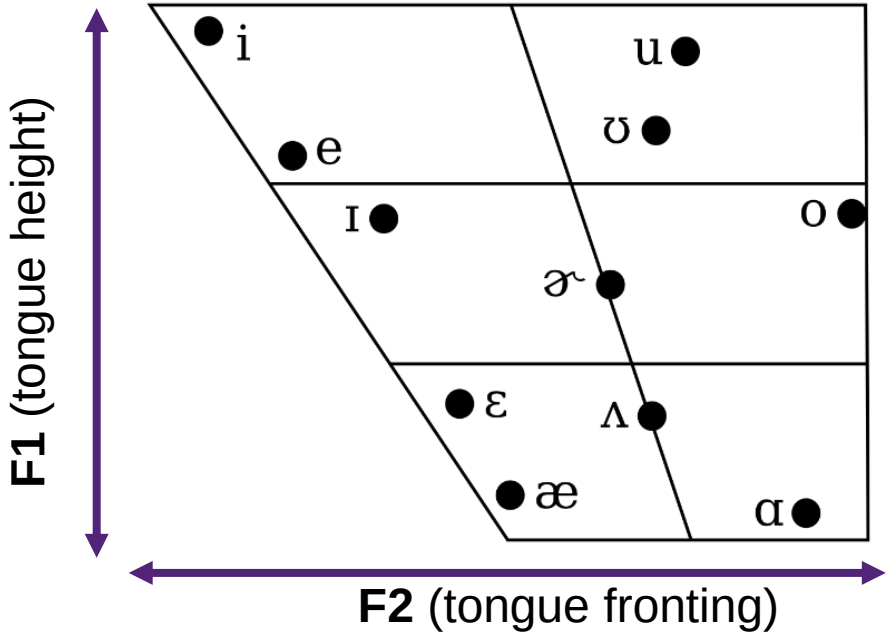
Background and Motivation



What are “formants”? and do they have to do with tongue position?

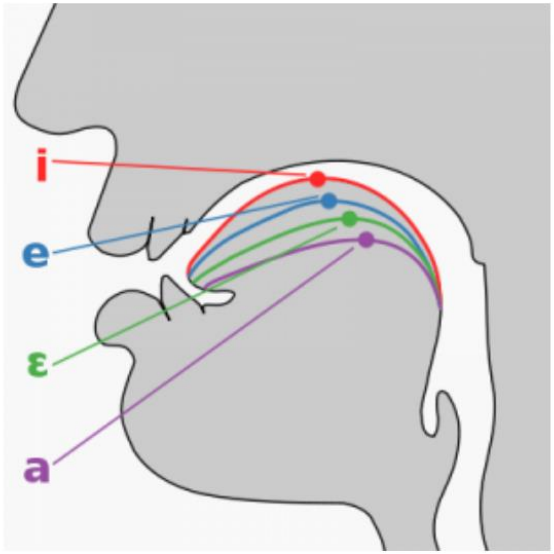
- Formants are concentration of acoustic energy at a certain frequency
  (Ladefoged & Johnson 2014)
- First formants (F1) inversely correspond to the tongue height
- Second formants (F2) and inversely correspond to tongue fronting





Left: Vowel chart of southern Californian Engl.

Research Gaps | Research Questions



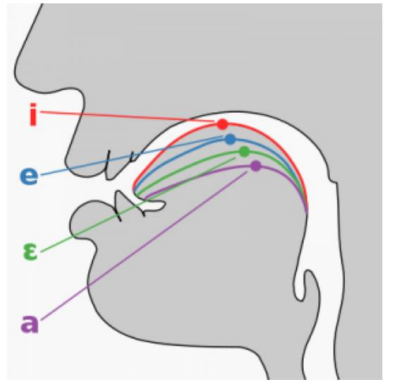
Problems and gaps in previous research

- Very little research
- Existing research very impressionistic (few systematic empirical research despite English being one of the largest learner groups globally!)
  - Learner vowel traits in naturalistic speech environments largely unknown
  - Limited generalisability | applicability of the findings
- **Current Study**
  - **Larger-scale analysis** of Chinese learners of English vowels
  - produced under more **spontaneous speech** conditions is needed!

RQ1: Do Chinese learners **merge /i:/ and /ɪ/ as well as /e/ and /æ/?**

RQ2: Do Chinese learners **show significant deviations from L1 speakers in terms of vowel duration for /u:/ and /ʊ/ as well as /i:/ and /ɪ/?**

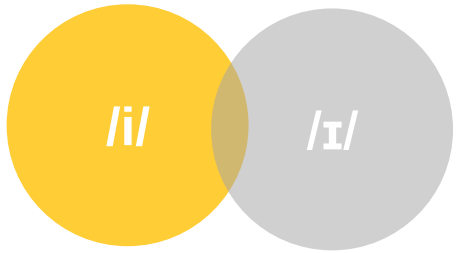
Research Gaps | Research Questions



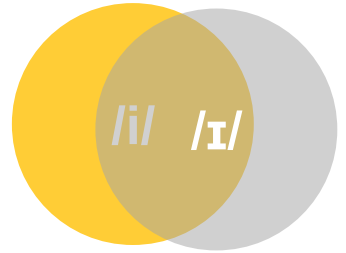
State of the art

- Japanese learners merge /i:/ and /ɪ/ (Zhang & Yin, 2009) and do not differentiate between /e/ and /æ/ (Wang, 2008; Zhang & Lu, 2012; Jiang & Zhang 2019)

L1 English speaker (ENS)



Chinese learner of English (CHN)

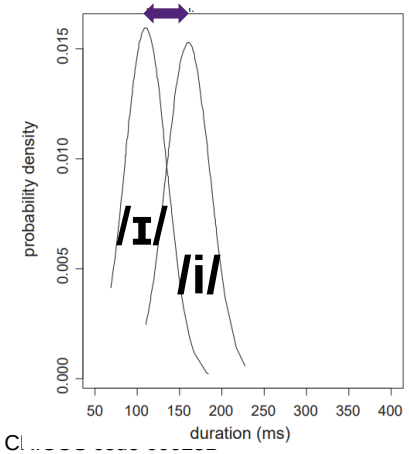


Hypothesis

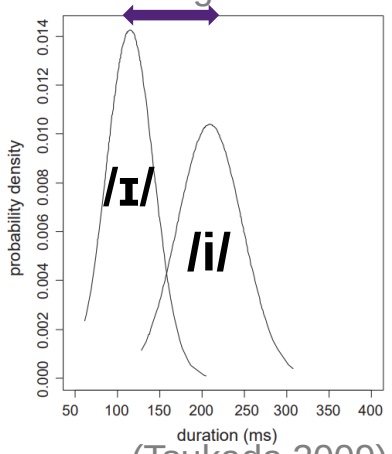
- Chinese speakers will show mergers for /i:/ and /ɪ/ and /e/ and /æ/

- Chinese speakers will show significant deviations in durations of vowels compared to L1 English speakers (similar to Japanese learners of English where Japanese also lack phonemic duration contrasts and thus exaggerate contrasts compared to L1 English speakers (Schweinberger & Komiya 2022).

L1 English speakers



Japanese learners of English



(Tsukada 2009)

Methodology (Data | Analysis)

Data

- *International Corpus Network of Asian Learners of English* (ICNALE) (Ishikawa 2014)
  - Speech and text samples from English learners in Asia and L1 English speakers
  - Spoken monologues: spontaneous speech from **148 Chinese learners** and **87 L1 speakers of English**
  - Final data set

Type	Speakers	/æ/	/ɛ/	/i/	/ɪ/	/u/	/ʊ/	Total
CHN	148	514	228	723	795	278	219	2,757
ENS	87	869	340	660	443	255	110	2,677
Total	**235**	**1,383**	**568**	**1,383**	**1,238**	**533**	**329**	**5,434**

Right: Gender and proficiency levels of CHN

Gender	A2	B1	B2	Total
female	9	70	7	86
male	5	55	2	62
Total	**14**	**125**	**9**	**148**

Methodology (Data | Analysis)

Data Processing (R 4.2.2, R Core team (2022) in RStudio (RStudio Team 2022))

- Aligning speech with audio using Web-MAUS (Schiel 1999)
  (this produces Praat TextGrids)
- Automated extraction of vowel formants and vowel duration from Praat TextGrids (Wickham et al. 2019)
- Only monosyllabic words were retained and outliers were removed using Kernel Density Estimation

Statistical Analysis

- Mergers → **Bhattacharya affinity** (Johnson 2015, measure of overlap of scatter clouds, 1 = perfect overlap)
- Duration → **Mixed-Effects Regression Model** (lme4: Bates et al. (2015), sjPlot: Lüdeke (2021))
  - DV: duration
  - IVs: type, vowel, gender, age, word type
  - REs: word, speaker

Results

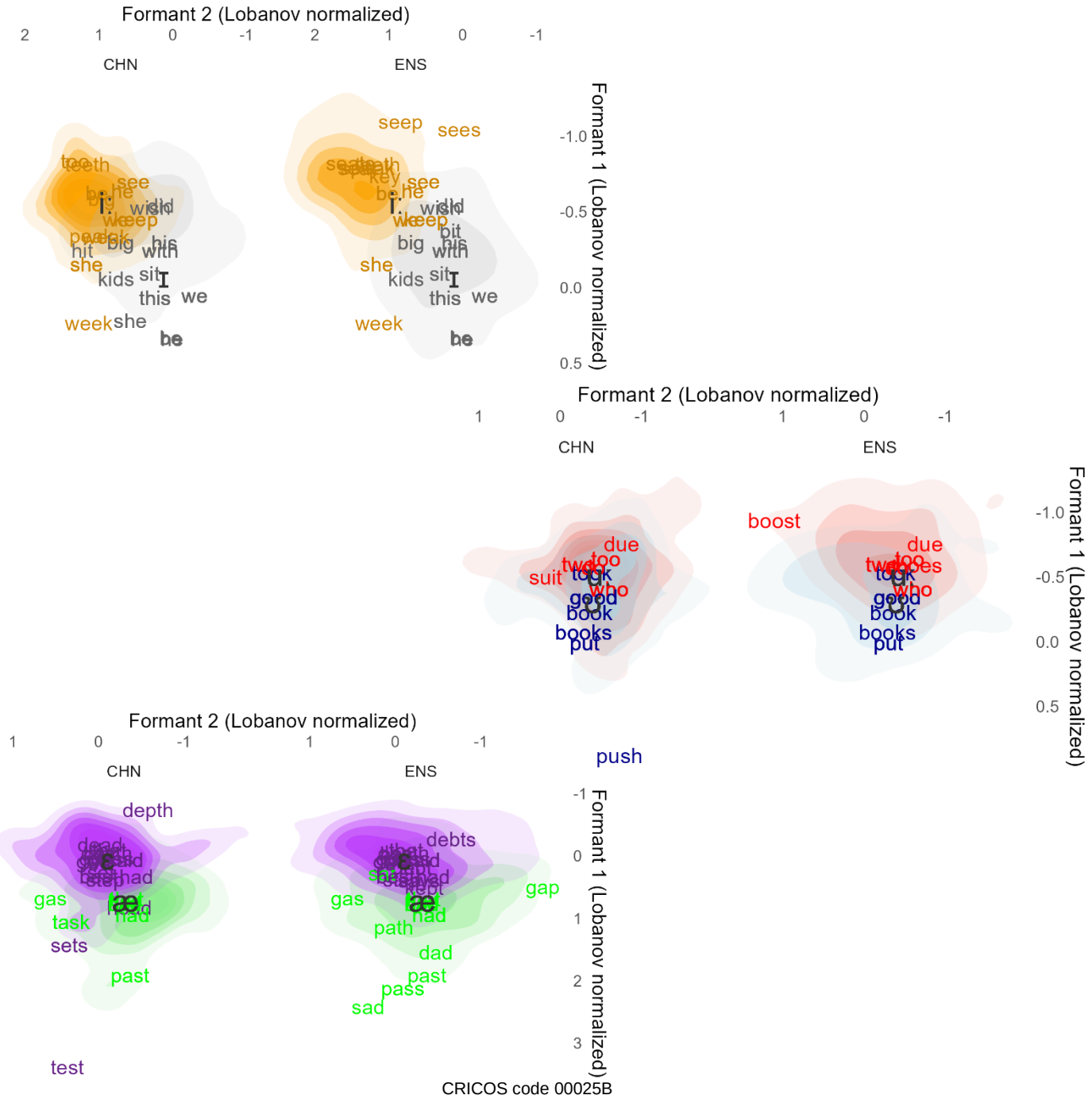
Mergers

- /i:/ and /ɪ/
  - CHN Bhattacharya affinity: .833
  - ENS Bhattacharya affinity: .692

  Substantively more overlap among CHN!
- /u:/ and /ʊ/
  - CHN Bhattacharya affinity: .939
  - ENS Bhattacharya affinity: .956

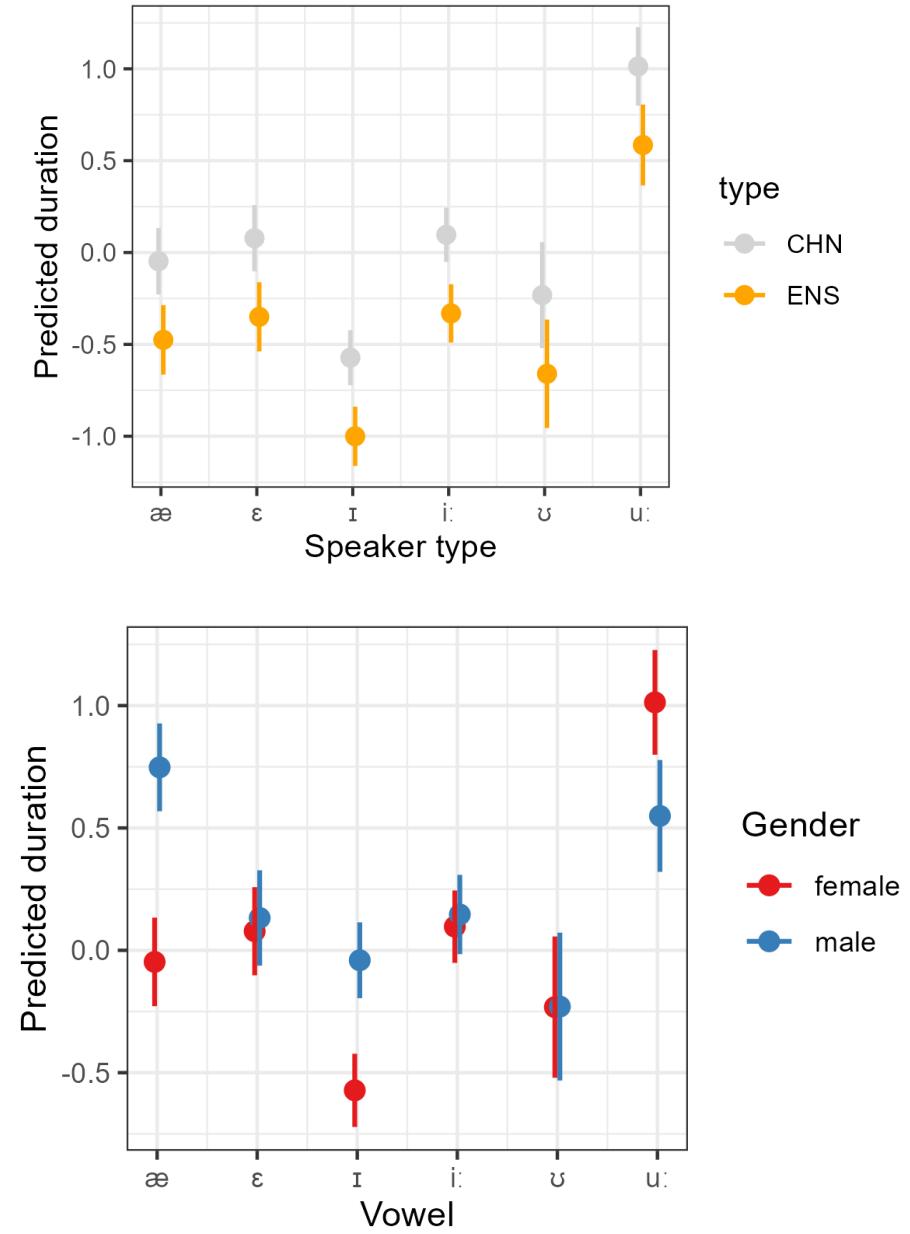
  Substantive overlap among both groups
- /æ/ and /ɛ/
  - CHN Bhattacharya affinity: .797
  - ENS Bhattacharya affinity: .816

  Overlap among both groups



Results

Final minal adequate model (N = 5,434)			
Predictors	**Estimates**	**CI**	**p**
(Intercept)	-0.05	-0.23 – 0.13	0.608
type [ENS]	-0.43	-0.51 – -0.34	<0.001
WordClass [lexical]	0.25	0.08 – 0.42	0.003
Vowel [ɛ]	0.13	-0.02 – 0.27	0.082
Vowel [ɪ]	-0.52	-0.73 – -0.32	<0.001
Vowel [iː]	0.14	-0.06 – 0.35	0.163
Vowel [ʊ]	-0.18	-0.48 – 0.11	0.227
Vowel [uː]	1.06	0.80 – 1.32	<0.001
Gender [male]	0.8	0.68 – 0.92	<0.001
Vowel [ɛ] × Gender [male]	-0.74	-0.92 – -0.56	<0.001
Vowel [ɪ] × Gender [male]	-0.26	-0.41 – -0.12	<0.001
Vowel [iː] × Gender [male]	-0.75	-0.89 – -0.60	<0.001
Vowel [ʊ] × Gender [male]	-0.79	-1.02 – -0.57	<0.001
Vowel [uː] × Gender [male]	-1.26	-1.45 – -1.07	<0.001
Random Effects / Model statistics			
τ00 Speaker (N)		0.06 (235)	
τ00 Word (N)		0.03 (37)	
Marginal R^2 / Conditional R^2		**0.179 / 0.263 (baseline 0.229)**	



Discussion | Outlook

Comparison with previous findings

- **Confirmation | Substantiation**
  - CHN: mergers of /i:/ and /ɪ/ (Zhang & Yin, 2009) as well as /e/ and /æ/ (Wang, 2008; Zhang & Lu, 2012; Jiang & Zhang 2019) (RQ1)
  - ENS: merger of /u:/ and /ʊ/ (Schweinberger & Komiya, 2022)
- **Unique findings | Conflicts**
  - CHN: merger also for /u:/ and /ʊ/ (RQ1)
  - CHN: produce all vowels shorter than ENS (RQ2)
  - CHN: do not exaggerate duration contrasts like Japanese learners of English (Schweinberger & Komiya, 2022; Tsukada 2009)
  - ENS: merger of /e/ and /æ/ in spon. speech

Limitations

- Did not substrate of CHN into account (Mandarin vs Cantonese vs etc.)
- Did not check for vowels in CHN substrate.
- Quality of recordings is really poor! (minute-long recordings recorded on cell phones)!

Significance

- Bad quality could | can be compensated using advanced methods (Kernel Density Estimation)
- Insights into vowel production by JPN learners in **spontaneous speech** (underexplored) → natural setting allows to **generalise findings to real-life learner speech**
- **Automated** corpus-based investigation on larger samples

Potential Applications

Prototype (proof-of-concept)

- Extend study to other learners of English
  (especially learners with L1's whose vowel system differs from English)

Significance

- **One of the first large-scale, corpus-based studies of ESL vowel production in natural speech!**
- Follow-up: perception → do differences in vowel production correspond to difficulties in intelligibility?

Possible Applications

- Creation of targeted classroom materials to improve L1-like vowel production among learners

Thank you really very much!

Ruihua Yin
Martin Schweinberger

ruihua.yin@uq.edu.au
m.schweinberger@uq.edu.au

References

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*, 233-277. York Press.

Grenon, I., Kubota, M., & Sheppard, C. (2019). The creation of a new vowel category by adult learners after adaptive phonetic training. *Journal of Phonetics* 72: 17-34.

Jiang, L., & Zhang, Y. (2019). Perception and production of English vowels by Chinese learners. Journal of Second Language Pronunciation, 5(1), 1-20.

Johnson, D. E. (2015). Quantifying Overlap with Bhattacharyya's affinity and other measures! Paper presented at NWAV15 (https://danielezrajohnson.shinyapps.io/nwav_44/).

Ladefoged, P., & Johnson, K. (2014). *A Course in Phonetics*. Cengage.

Schweinberger, M., & Komiya, Y. (2022). A corpus-based computational analysis of high-front and -back vowel production of L1-Japanese learners of English and L1-English speakers. In Rosey Billington (ed.), *Proceedings of the Eighteenth Australasian International Conference on Speech Science and Technology*, 196-200. Australasian Speech Science and Technology Association.

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)*, 305-308.

Wang, X. (2008). Pronunciation problems of Chinese students learning English: Interference and instruction. *US-China Education Review*, 5(9), 16-22.

Wang, W. S., & Sun, C. (2015). *The Oxford handbook of Chinese linguistics*. Oxford University Press.

Zhang, F. & Yin, P. (2009). A Study of Pronunciation Problems of English Learners in China. *Asian Social Science*, 5(6), 141-146.

Zhang, Y., & Lu, Y. (2012). A study on pronunciation problems of English learners from China. *English Language Teaching*, 5(12), 156-162.

Data & Software

Data

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.

Software

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

https://github.com/MartinSchweinberger/ICAME44

RStudio Team (2022). *RStudio: Integrated Development Environment for R*. RStudio, PBC. Boston, MA (http://www.rstudio.com)

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 (https://CRAN.R-project.org/package=sjPlot).

Schiel, F. (1999). Automatic Phonetic Transcription of Non-Prompted Speech, *Proceedings of the ICPhS 1999. San Francisco, August 1999*, 607-610.

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

A corpus-based acoustic analysis of vowel production by L1-Chinese learners and native speakers of English

Martin Schweinberger
Ruihua Yin

m.schweinberger@uq.edu.au
ruihua.yin@uq.edu.au

All materials (except data) available at https://github.com/MartinSchweinberger/ICAME44