THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
CREATE CHANGE

atap australian text analytics platform
atap.edu.au

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

AcqVA Aurora Lab

Yuki Komiya
Martin Schweinberger

LADAL

y.komiya@uq.edu.au
m.Schweinberger@uq.edu.au

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

Timeline | Table of Contents

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

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.)

Pronunciation is important for learners as well as teachers of English!

Background and Motivation

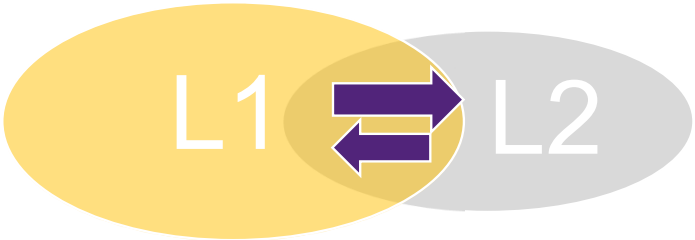
Why is pronunciation a challenge for L2 English learners?

Languages interact in the minds of multilingual speakers

Languages are not independent but affect each other

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)

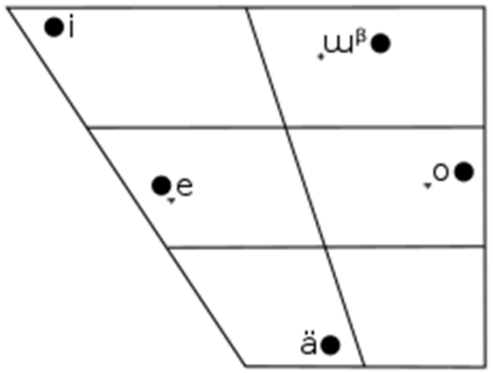


Background and Motivation

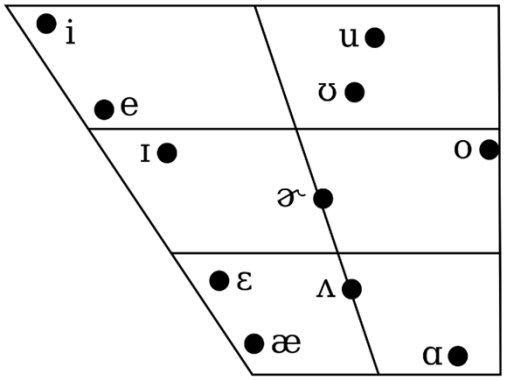
English vowels are particularly challenging for Japanese-L1 learners
(Franklin & Stoel-Gammon 2014)

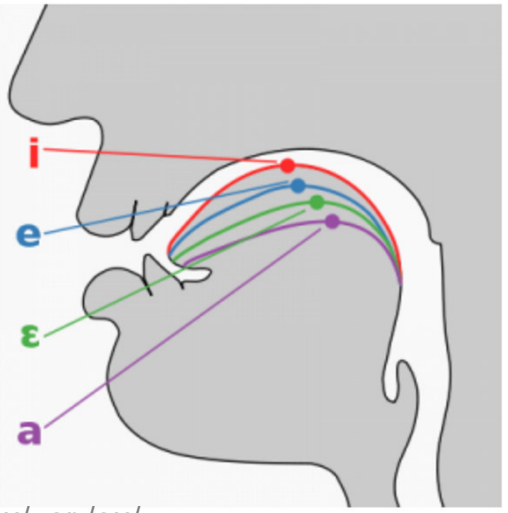
- Differences in inventory size (JPN: 5 vowels* vs. ENS: app.** 11 vowels) (Homma 1992)
- Differences in how vowels are differentiated (ENS: formants + duration | JPN: duration)

Japanese vowel chart



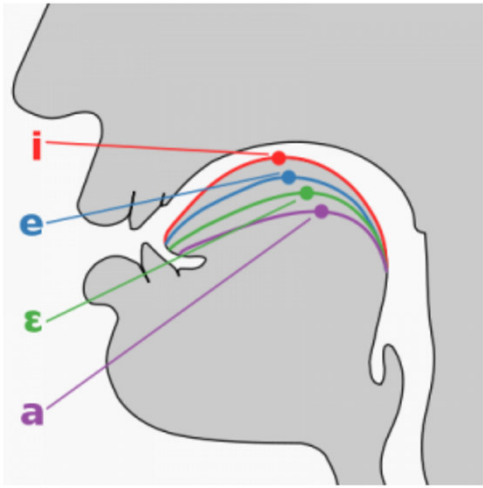
Vowel chart of
southern Californian Engl.





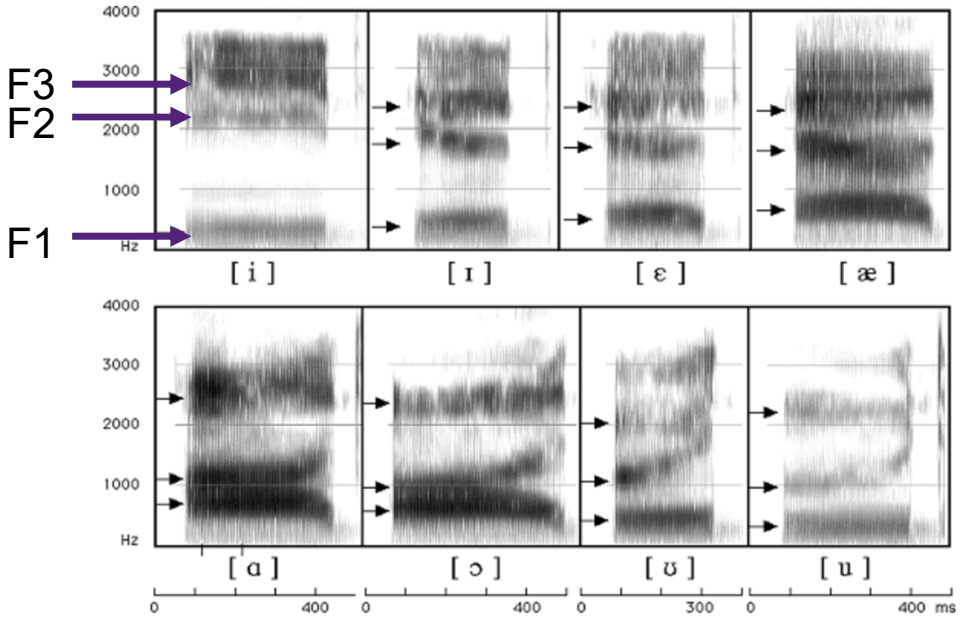
* Monophthingal vowels, i.e., vowels with a fixed tongue position (not vowels with a moving tonue position like /au/ or /ou/.
** Depending on the variety of English.

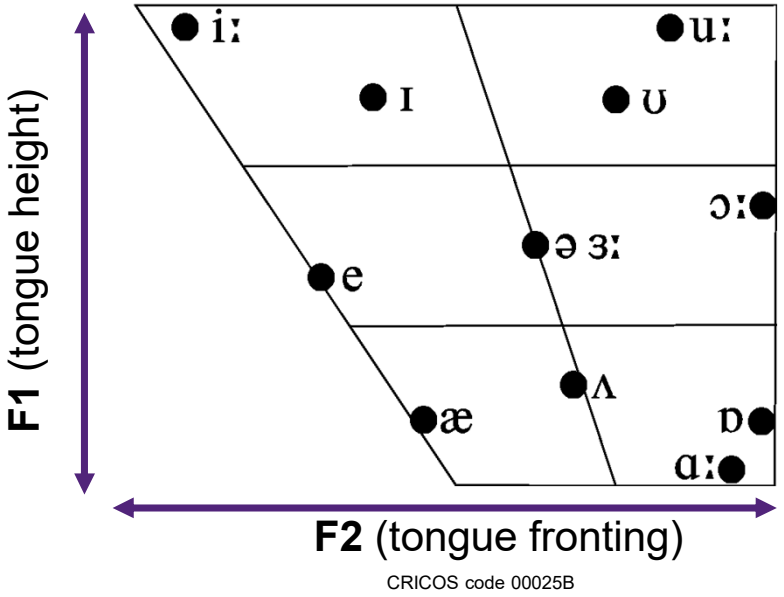
Background and Motivation



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) and second formants (F2) inversely correspond to the tongue height (F1) and tongue fronting (F2: where it is raised) of each vowel

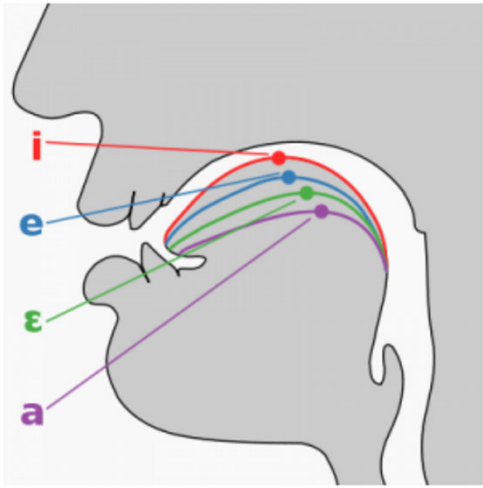




Left: Vowel chart of Received Pronunciation (RP)

CRICOS code 00025B

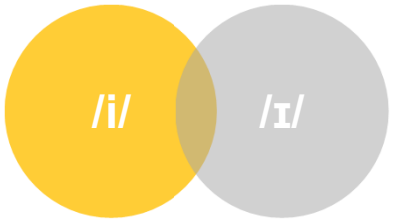
Research Gaps | Research Questions



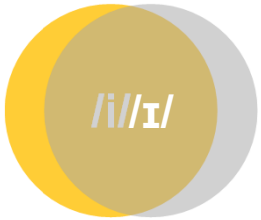
What has been said about English vowels produced by Japanese learners?

- Japanese learners merge spectrally similar vowels (Ingram & Park 1997, Ueyama 2003)

L1 English speaker (ENS)

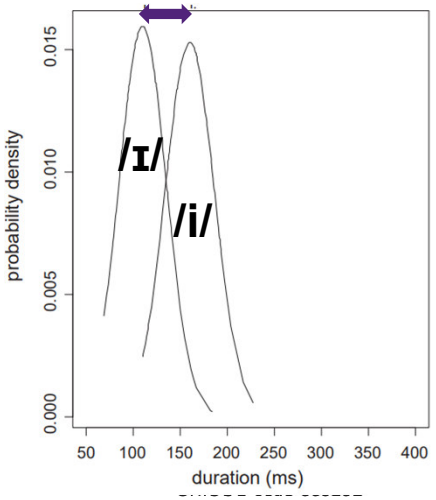


Japanese learner of English (JPN)

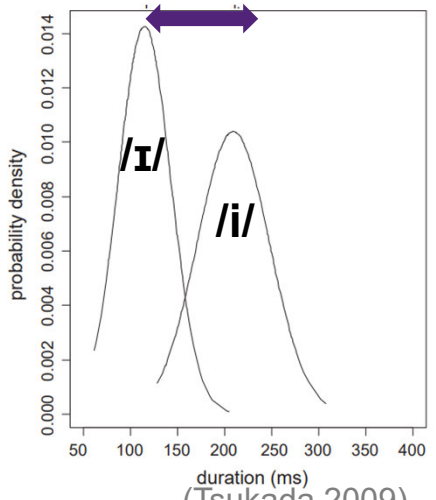


- Japanese speakers are **very sensitive** to vowel duration (Kato et al. 2001) and **exaggerate duration to compensate** for the relative insensitivity to formant differences (Morrison 2002)

L1 English speakers

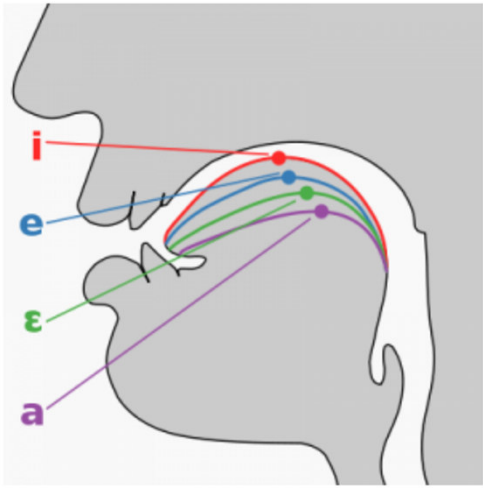


Japanese learners of English



(Tsukada 2009)

Research Gaps | Research Questions



Problems and gaps in previous research

- Investigation mainly done in highly controlled laboratory conditions (scripted word | sentence-reading)
  - ➢ Learner vowel traits in naturalistic speech environments largely unknown
- Small subject size (±10 speakers)
  - ➢ Limited generalisability | applicability of the findings

Larger-scale analysis of vowels
produced under more **spontaneous speech** conditions is needed!

RQ1: Do Japanese learners **merge /i:/ and /ɪ/ as well as /u:/ and /ʊ/**?

RQ2: Do Japanese learners **exaggerate the length of vowels** to compensate lack of spectral differentiation?

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 **150 Japanese learners** and **132 L1 speakers of English**
  - Before processing

Type	Speakers	/ɪ/	/i:/	/ʊ/	/u:/
ENS	132	2,562	1,205	350	1,895
JPN	150	3,696	1,203	644	1,261
Total	**282**	**6,258**	**2,408**	**994**	**3,156**

  - After processing (final data set)

Type	**Speakers**	**/ɪ/**	**/i:/**	**/ʊ/**	**/u:/**
ENS	105	693	939	189	395
JPN	141	1,122	535	188	281
Total	**246**	**1,815**	**1,474**	**377**	**676**

Methodology (Data | Analysis)

Data Processing (R 4.2, 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

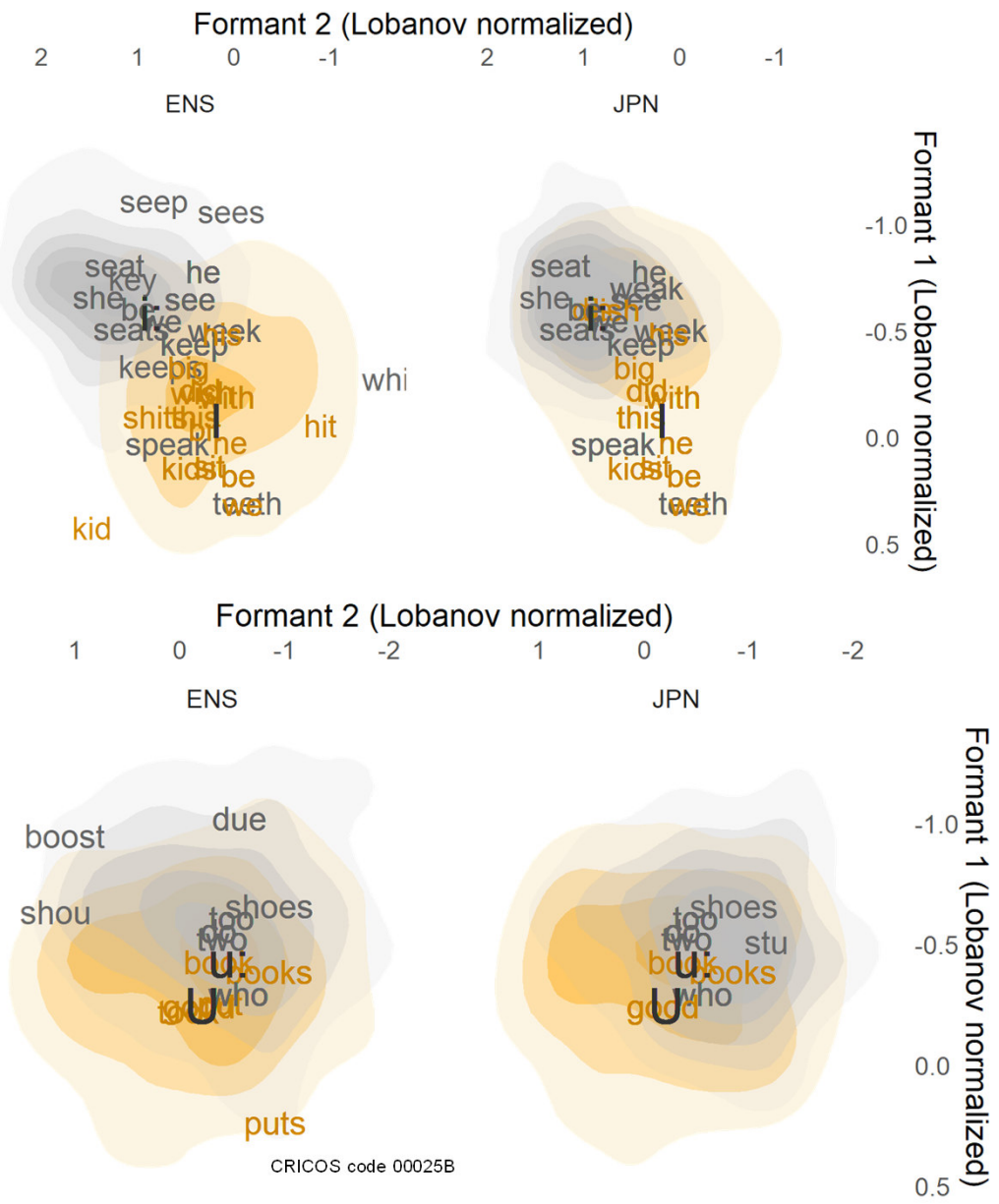
Results

Mergers

- /i:/ and /ɪ/
  - JPN Bhattacharya affinity: .901
  - ENS Bhattacharya affinity: .757

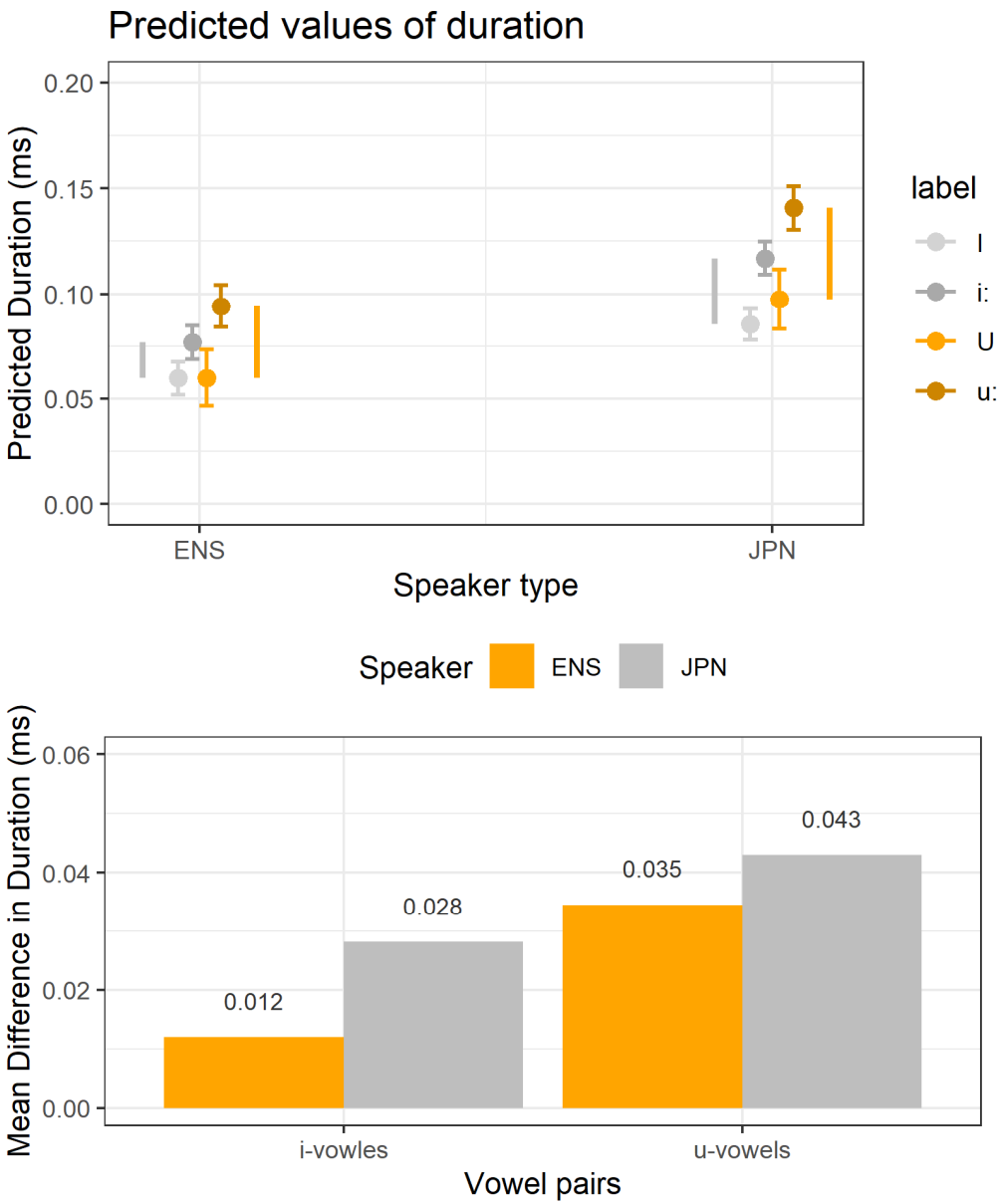
  Substantively more overlap among JPN!
- /u:/ and /ʊ/
  - JPN Bhattacharya affinity: .932
  - ENS Bhattacharya affinity: .952

Mergers confirmed for **spectrally similar vowels** (ENS also merge /u:/ and /ʊ/)!



Results

	final minimal adequate model		
Predictors	Est.	CI	p
(Intercept)	0.06	0.05 - 0.07	<0.001
type_{JPN}	0.03	0.02, 0.03	<0.001
$\text{label}_{i:}$	0.02	0.01, 0.02	<0.001
label_{U}	0.00	-0.01, 0.01	0.981
$\text{label}_{U:}$	0.03	0.02, 0.05	<0.001
gender_{male}	0.01	0.00, 0.02	0.001
type_{JPN} * $\text{label}_{i:}$	0.01	0.01, 0.02	0.001
type_{JPN} * label_{U}	0.01	-0.00, 0.02	0.056
type_{JPN} * $\text{label}_{U:}$	0.02	0.01, 0.03	<0.001
type_{JPN} * gender_{male}	-0.02	-0.03, -0.01	<0.001
Random Effects			
ICC		0.07	
N		149 speaker 46 word	
Observations		4342	
Mar. R^2 / Cond. R^2		0.115 / 0.178	

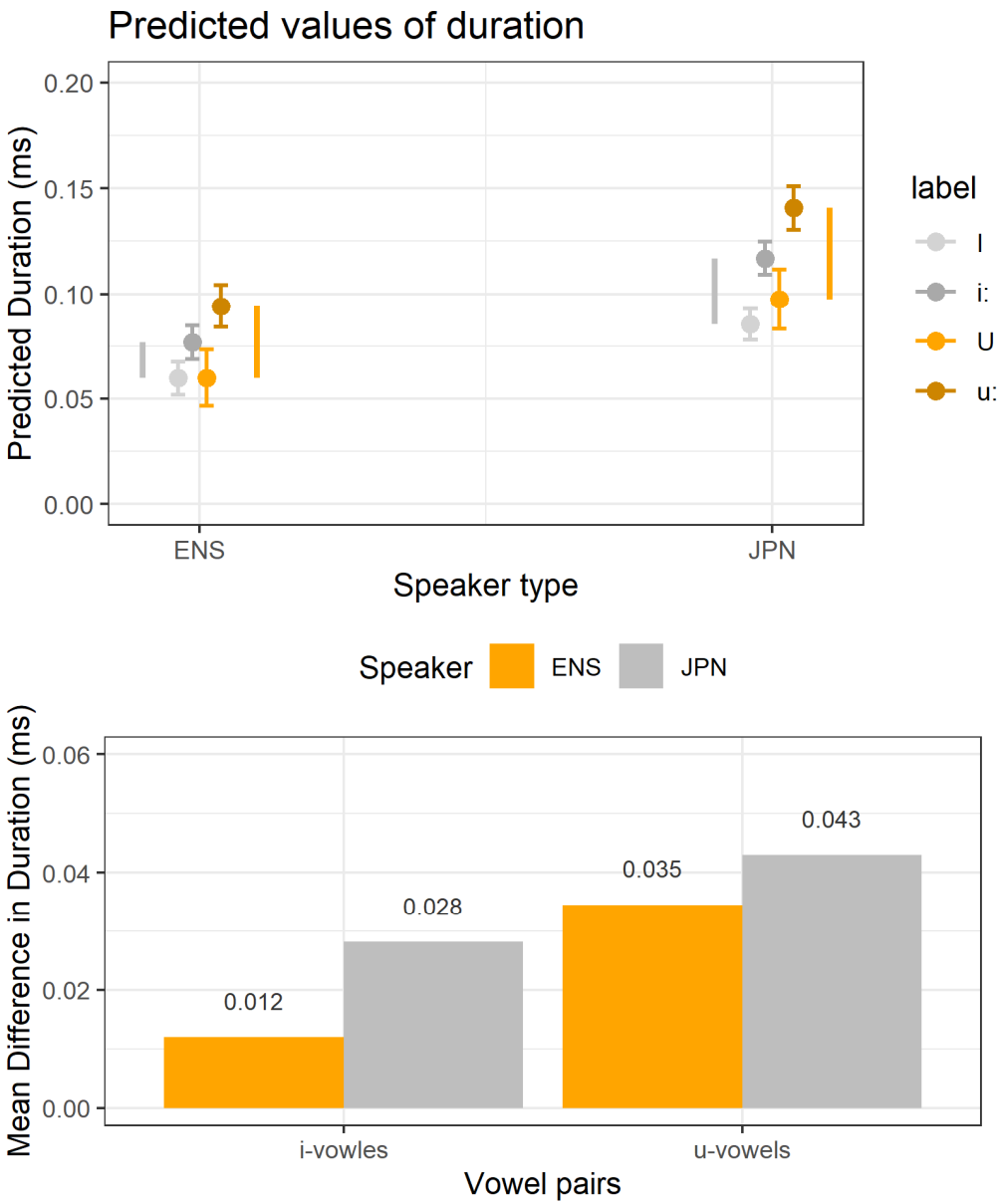


Results

Duration

- JPN extend all vowel durations (not just long vowels)! (expectation: short vowels shorter | long vowels longer!)
- JPN exaggerate the duration difference of both /i:/ and /ɪ/ as well as /u:/ and /ʊ/

Exaggerated duration difference by JPN speakers confirmed for **both /i:/** and **/ɪ/** as well as **/u:/** and **/ʊ/!**



Discussion | Outlook

Comparison with previous findings

- **Confirmation | Substantiation**
  - JPN: mergers of spectrally close vowels (lab settings: Ueyama 2003; Tsukada 2001)
- **Unique findings | Conflicts**
  - ENS: merger of /u:/ and /ʊ/ in spon. speech
  - JPN: exaggerated durational contrasts between spectrally similar vowels in nat. settings (lab settings: Tsukada 2009)

1. Apply same method to German learners and learners of other languages (e.g. German)
2. Determine what factors differentiate ENS and L2 speakers re. vowel production

(potential MA theses!)

Limitations

- Quality of recordings is really poor! (minute-long recordings recorded on cell phones)!
- Difficult to control semantic | phonological environments (which is important) (see Visceglia et al. 2009)

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 L1 German learners of English (or learners of other languages)

Significance

- **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?

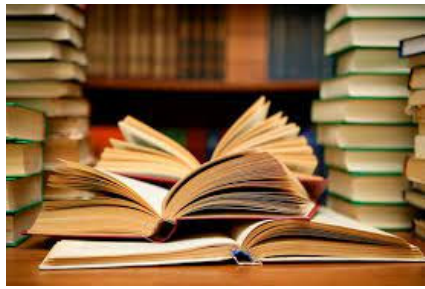
Possible Applications

- Creation of targeted classroom materials to improve L1-like vowel production among learners
- Convert analysis into a mobile app for MELLT (Mobile-Enhanced Language Learning and Teaching)
- BMBF | Volkswagen | ERC grant proposal on ***Improving Language Production among Language Learners via Direct Digital Feedback*** (Collaboration with the Phonetics group in the ISFAS at Uni Kiel)

Thank you really very much!

Yuki Komiya
Martin Schweinberger

y.komiya@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.

Franklin, A. D., & Stoel-Gammon, C. (2014). Using multiple measures to document change in English vowels produced by Japanese, Korean, and Spanish speakers: The case for goodness and intelligibility. *American Journal of Speech-Language Pathology* 23(4): 625-640.

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.

Homma, Y. (1992). *Acoustic phonetics in English & Japanese*. Yamaguchi Shoten.

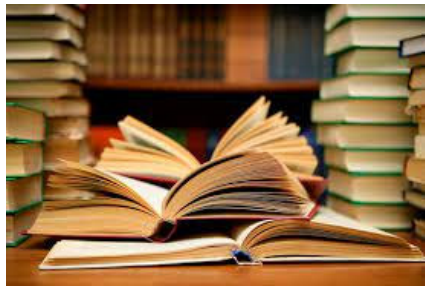
Ingram, J. C. L., & Park, S. G. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics* 25(3): 343-370.

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

Kato, H., Tajima, K., & Akahane-Yamada, R. (2001). Native and non-native perception of phonemic length contrasts in Japanese. *The Journal of the Acoustical Society of America* 110(5): 2686.

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

Morrison, G. S. (2002). Japanese listeners' use of duration cues in the identification of English high front vowels. In J. Larson. & M. Paster (eds.), *Proceedings of the 28th annual meeting of the Berkeley Linguistics Society*, 189–200. Berkeley Linguistics Society.



References

Satoi, H., Yoshimura, M., & Yabuuchi, S. (2005). The relationship between English speech rhythm and vowel reduction in production: Comparison between Japanese EFL learners and native English speakers. *Language Education & Technology* 42: 59-72.

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.

Tsukada, K. (2009). Durational characteristics of English vowels produced by Japanese and Thai second language (L2) learners. *Australian Journal of Linguistics* 29(2): 287-299.

Ueyama, M. (2003). Duration and quality in the production of the vowel length contrast in L2 English and L2 Japanese. In M. J. Solé., D. Recasens., & J. Romero. (eds.), *15th International Congress of Phonetic Sciences*, 1509-1512. Universitat Autònoma de Barcelona.

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

Yamaguchi, T., & Chiew. P. S. (2020). Is there conflation? An acoustic analysis of vowels in Japanese English. *Asian Englishes* 22(1): 35-51.

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)

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.

Acknowledgements

The study is part of a project that Yuki Komiya and Martin Schweinberger worked on at the University of Queensland.

This paper has been submitted to *SST2022 - The 18th Australasian International Conference on Speech Science and Technology*.

Yuki Komiya
y.komiya@uq.edu.au
https://www.linkedin.com/in/yuki-komiya-117004200/

Martin Schweinberger
m.schweinberger@uq.edu.au
martin.schweinberger@uit.no
https://github.com/MartinSchweinberger

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

Yuki Komiya
Martin Schweinberger

y.komiya@uq.edu.au
m.Schweinberger@uq.edu.au