Automatic Prediction of Pronunciation Errors by Second Language Learners based on Phonological and Phonetic Information of Learners' First Language and the Target Language

Shuju Shi

ECE590 SIP

March 17 , 2021

Overview

- Research Background
- Research Questions
- Methodology
  - Stimuli design and corpus
  - Feature extraction and normalization
  - Assimilation of perceptual space using acoustic features
- Experiments and Results
  - Acoustic analysis of vowel inventories
  - Assimilation of L2 pronunciation
- Conclusion and Future Work

Research Background

- The phenomenon:
  - Learners' L1 has a systematic influence on their L2 sound acquisition
- The theories:
  - Perceptual Assimilation Model (PAM/PAM-L2)
  - Speech Learning Model (SLM/SLM-r)
  - Native Magnet Theory Model (NLM)
- The applications:
  - Simulating L1/L2 perceptual space (Guenther and Gjaja 1996, Shi and Shih 2019)
  - Simulating L2 sound acquisition process (Thomson et al. 2009, Gong et al. 2015)

Perceptual Assimilation Model (Best 1995, Best & Tyler 2007)

- PAM accounts for how naïve speakers (PAM) and L2 learners (PAM-L2) assimilate a new sound contrast in L2 according to their L1 phonology categories.

Table 1. The PAM-L2 assimilation patterns for non-native contrasts. (Adapted from Best 1995, pp. 125)

Category	**Assimilation Pattern**	**Prediction**
Two-Category	Two L2 Sounds → Two L1 sounds	Excellent
Single-Category	Two L2 Sounds → One L1 sound	Poor
Category-Goodness	Two L2 sounds → One L1 sound	Variable (Poor to very good)
No L1-L2 Assimilation	Two L2 sounds → No L1 sound	New category/categories

Speech Learning Model and the revised Speech Learning Model (Flege, 1995 & 2021)

- SLM and SLM-r accounts for the variation in the extent of individuals' learning phonetic segments in an L2.
  - In contrast to PAM/PAM-L2, L1 and L2 are related perceptually at allophonic level.
  - Possibility of forming new L2 categories increases with perceived dissimilarity.
  - L2 sound categories may differ from the native categories.
  - Learners' ability to discern phonetic difference between L2 sounds that are non-contrastive in their L1 decreases as age of learning increases.

Native Magnet Theory Model (Kuhl 1992 & 2000, Iverson et al. 2003)

- NLM accounts for how L1 experience serves as language-specific filters to warp the acoustic dimensions and influence how sounds in L2 are perceived, i.e., the perceptual magnetic effect:
  - Decreasing perceptual sensitivity within a category and increasing sensitivity between categories
  - Facilitating perceptual sensitivity of native phonetic categories whereas inhibiting perceptual sensitivity of phonetic categories in foreign languages

Computational Approaches

- Simulation magnetic effect in L1/L2 perception
  - Guenther and Gjaja (1996) proposed to use a self-organizing neural network to simulate perceptual magnetic effect.
- Simulating L2 sound acquisition
  - Thomson et al. (2009) used discriminant function analysis to measure the similarity between Chinese and English vowels and then predict L2 learner behavior based on the achieved similarity degree.
  - Gong et al. (2015) introduced a framework where they used HMMs to model the interaction between L1 and L2 at the onset of L2 acquisition based on data of Chinese learner's perception of Spanish consonants.

Comparison of the theoretical models

- Common ground
  - All agree that L1 and L2 share a common phonological/phonetic space.
  - All establish their arguments based on the similarity/dissimilarity of sounds between L1 and L2.
- Features
  - SLM/SLM-r: perceived salient phonetic difference, distribution of sounds
  - PAM/PAM-L2: articulatory gestures
  - NLM: acoustic features
- Potential Problems
  - Features used in the first two models are more descriptive than quantitative.
  - Methods used in the third model are exhaustive and could be difficult if not impossible to implement on language-inventory level

Limitations of current computational models

- Stimuli
  - Synthetic speech
- Coverage of sound inventory
  - Subsets of either vowels or consonants of a language
- Assumption of assimilation level
  - Phonemic
- Pedagogical implications
  - Corrective feedback

Research questions

- Do L1 and L2 sound inventories exist in a common phonological/phonetic space?
- At what level (phoneme, allophone, orthography or a hybrid of the three) does L1 interfere L2 phonology/phonetics acquisition?
- How well can the quantified differences between L1 and L2 sound inventories account for L2 pronunciation errors?

Methodology: Phonological Vowel Inventories

Table 1. Mandarin Inventory: Orthographies

	FRONT	CENTRAL	BACK
HIGH	i ü		u
MID		e er	o
LOW	a		

Table 2. Mandarin Inventory: Phonemes

	FRONT	CENTRAL	BACK
HIGH	/i/ /y/		/u/
MID		/ə/ (/ɚ/)	(/ɤ/)
LOW	/a/		

Table 3. Mandarin Inventory: Allophones

	FRONT	CENTRAL	BACK
HIGH	[i] [y]	([ɻ] [ʅ])	[u] [ʊ]
MID	[e] [ɛ]	[ɚ] [ə]	[ɤ] [o]
LOW	[æ] [a] [aˠ]	[ɐ] [A]	[ɑ]

Table 4. English Inventory: Phonemes

	FRONT	CENTRAL	BACK
HIGH	/i/ /ɪ/		/u/ /ʊ/
MID	/ɛ/	/ɚ/ /ə/	/ʌ/ /ɔ/
LOW	/æ/		/ɑ/

Mandarin Diphthongs: /ai, au, ou, ei/

English Diphthongs: /aɪ, aʊ, oʊ, eɪ, ɔɪ/

Methodology: Stimuli Design and Corpus

- Participants
  - Chinese: 18 speakers (9 female, 9 male), Mandarin speakers, born and raised in Beijing, ages 19-34 (mean: 24.2, std.: 3.98), ages of English learning (6-10)
  - English: 13 speakers (7 male, 6 female), born and raised in the Chicago area, ages: 19-28 (mean:21.5, std.: 2.99)
- Stimuli
  - Chinese: all possible Chinese monosyllabic Pinyin with 4-tone variation (1856 syllables)
  - English: monosyllabic words selected based on frequency and of comparable size with the Chinese stimuli (1660 words)(COCA2016)
  - English speakers only do the recording for the English stimuli whereas Mandarin speakers do the recording for both English and Chinese.
  - Segmentation
    - Forced alignment: Montreal Forced Aligner
    - Manual checking

Methodology: Feature Extraction and Normalization

- The procedure to optimize formant ceiling follows the idea in Escudero et al. (2009).
  - Unit
    - Mandarin: tri-phone
    - English: 4 bi-phone conditions (V-/l/, V-/ɹ/, V-nasal, V-other)
  - Criteria
    - The “optimal ceiling” is chosen as the one that yields the lowest variation in the measured F1-F2 pairs among all the samples of that triphone/bi-phone.
- For each vowel, formants are extracted at its optimal ceiling, converted to bark, and then z-normalized within each speaker

Methodology: Perceptual Space Simulation

- Findings/Statements by the aforementioned theoretical models
  - L1 and L2 sounds share a common phonological/phonetic space
  - SLM: learners' representation of phonetic categories is based on different features, or feature weights, than native speakers’
  - NLM: A language learner's perceptual space of L2 sound inventory is distorted by his/her L1 sound inventory (Iverson et al., 2003)

PCA could possibly be used to address all the conditions.

Methodology: Perceptual Space Simulation (cont'd)

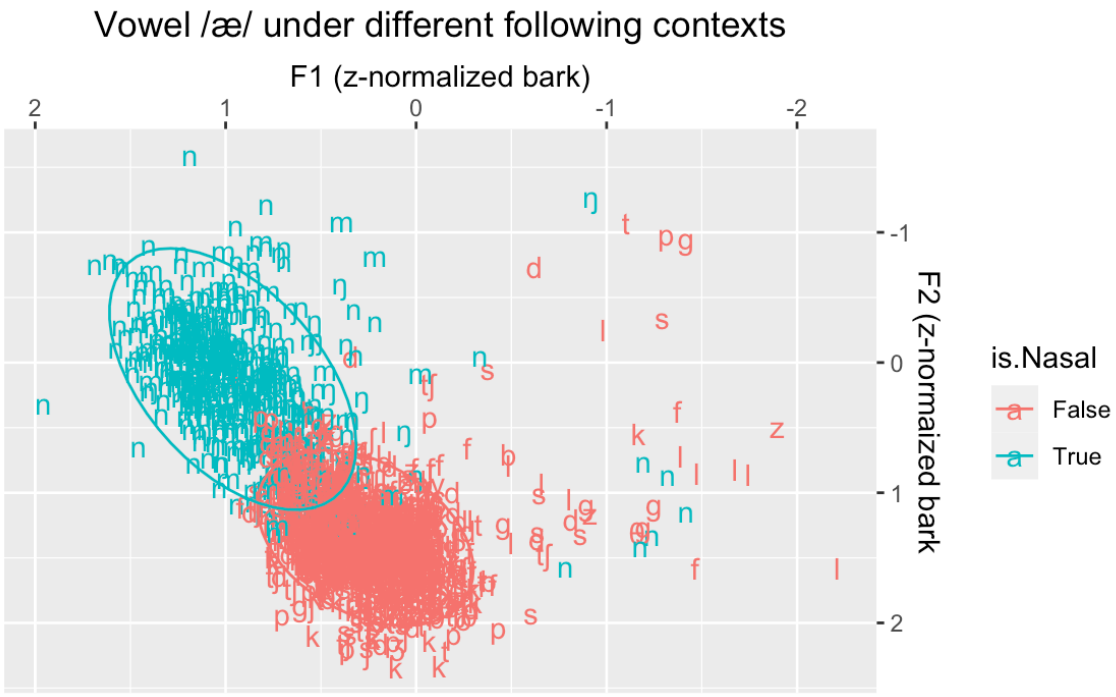
- In this study we proposed to use PCA in three different ways regarding how we get the principal components:

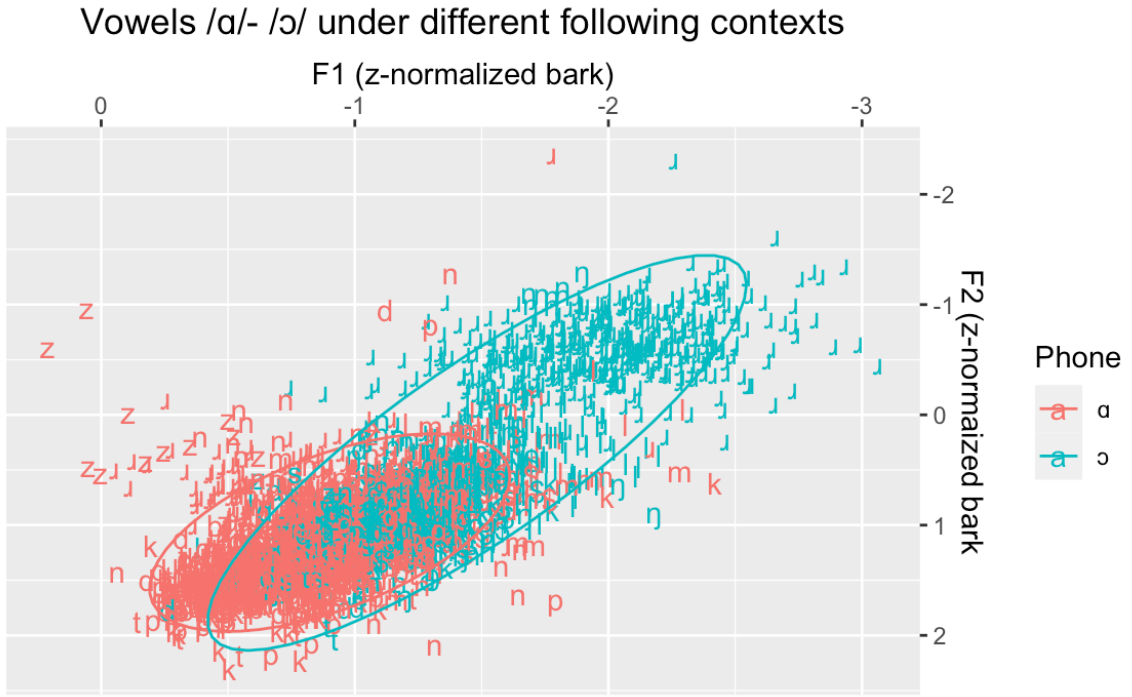
Table 5. Assumptions for the proposed PCA approaches

	Phonological Space	**Feature/Feature weights**
PCA1 ($W=W_{L1}$)	Separate(?)	L1
PCA2 ($W=W_{Target}$)	Separate	Target Language
PCA3 ($W=W_{L1+Target}$)	Common	Combined

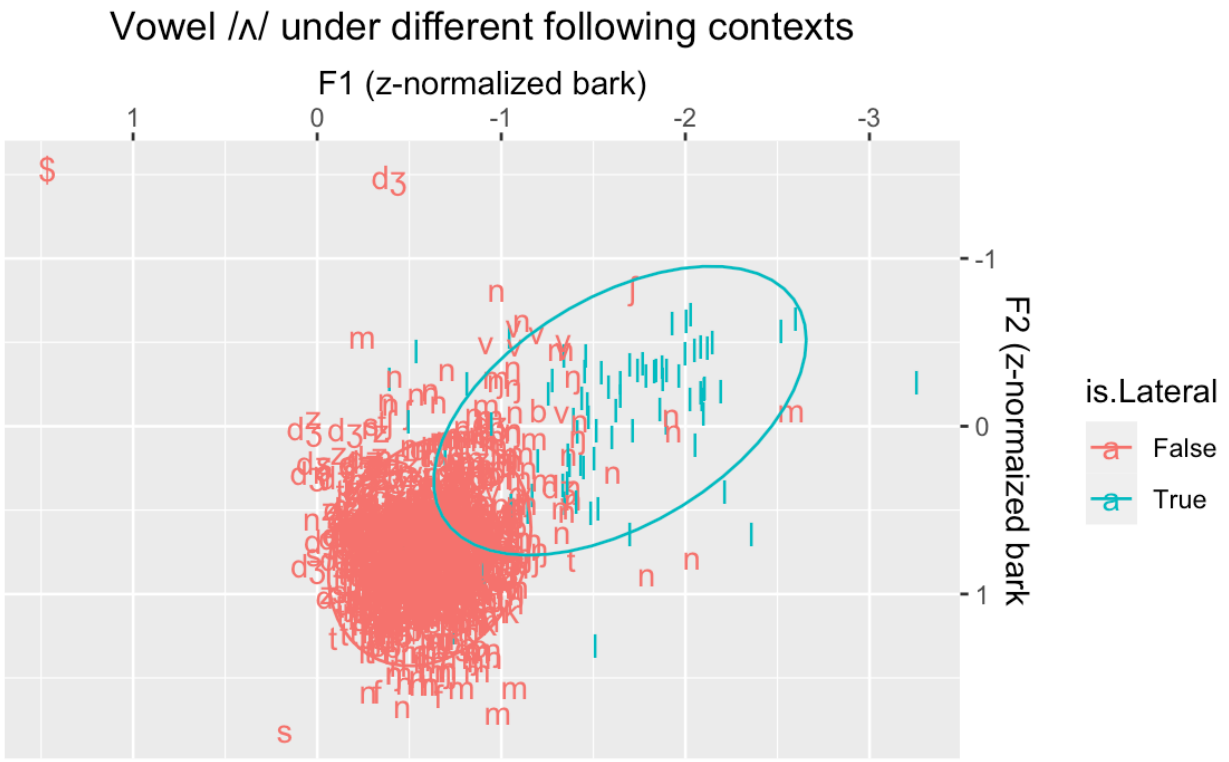
English Allophones

- Two allophones are included for /æ/: æ_nasal, æ_oral
- Three allophones are included for /ɑ/ and /ɔ/: ɑ_ɹ , ɔ_ɹ and ɑ-ɔ





English Allophones



- Two allophones are included for /ʌ/: ʌ_l and ʌ.
- In total, we end up with 18 vowels for English vowel allophone inventory.

Vowel classification at different inventory levels

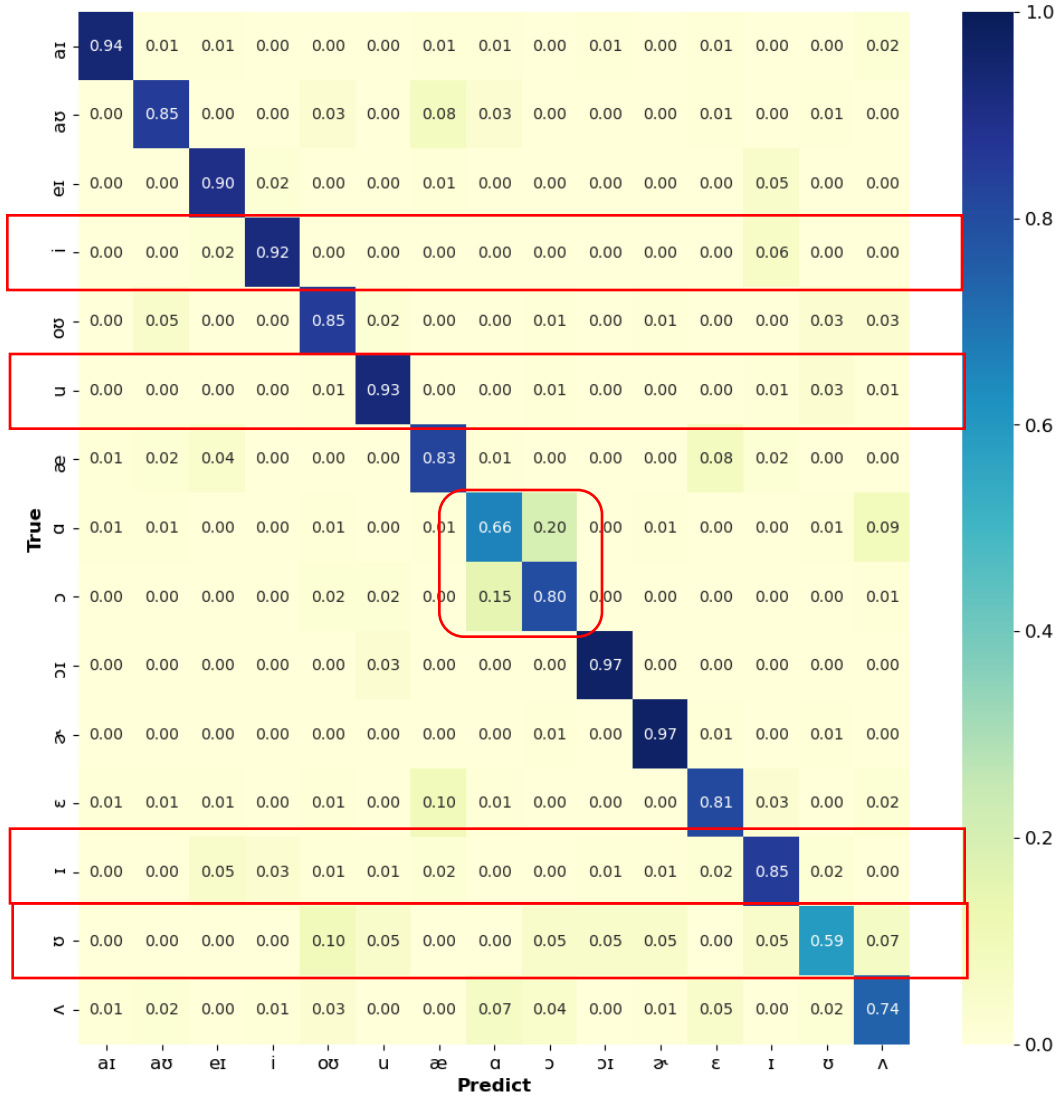
- Features
  - Duration, F1-F3 at 10 equally distributed time points of a vowel interval
- Model
  - Gaussian Mixture Models (GMMs)
- Results

Table 6. Classification results for vowels

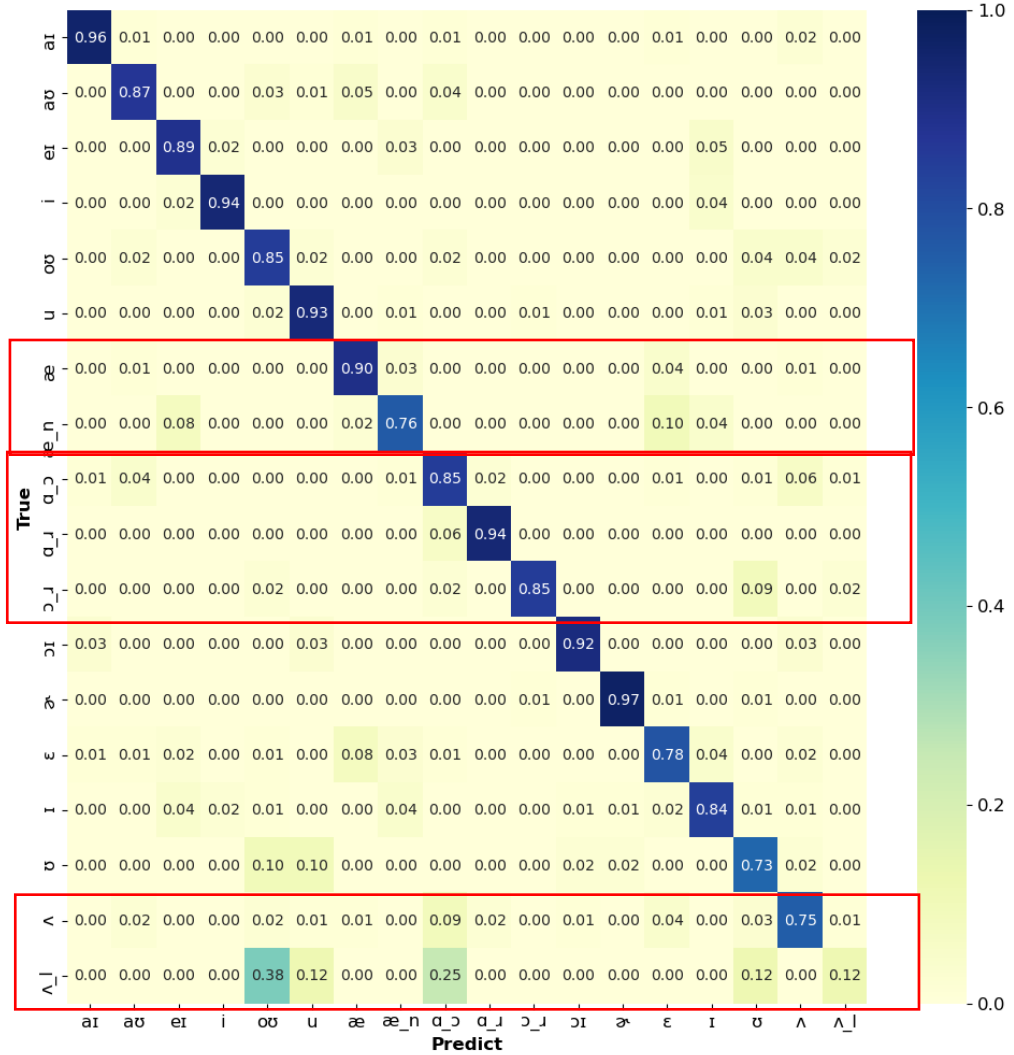
	English		**Mandarin**				
Level	Phoneme (15)	Allophone (18)	Pinyin (11)	Phoneme (10)	Phoneme (11)	Allophone (18)	Allophone (22)
Accuracy	86.3%	87.6%	90.1%	90.5%	90.8%	89.1%	87.5%

Classification results: English

Phoneme(15): 86.3%

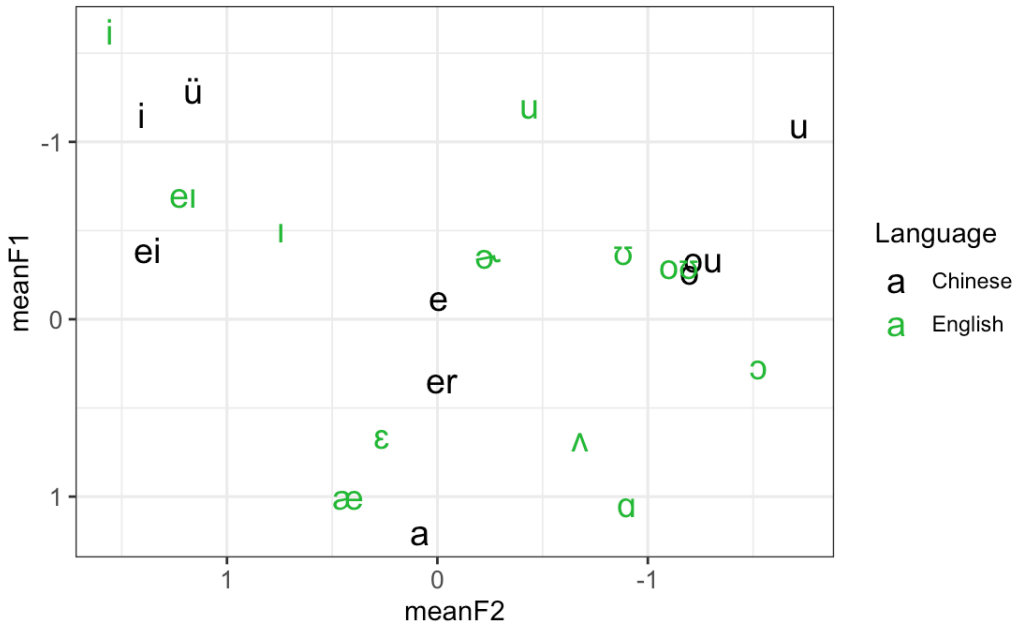


Allophone(18): 87.6%

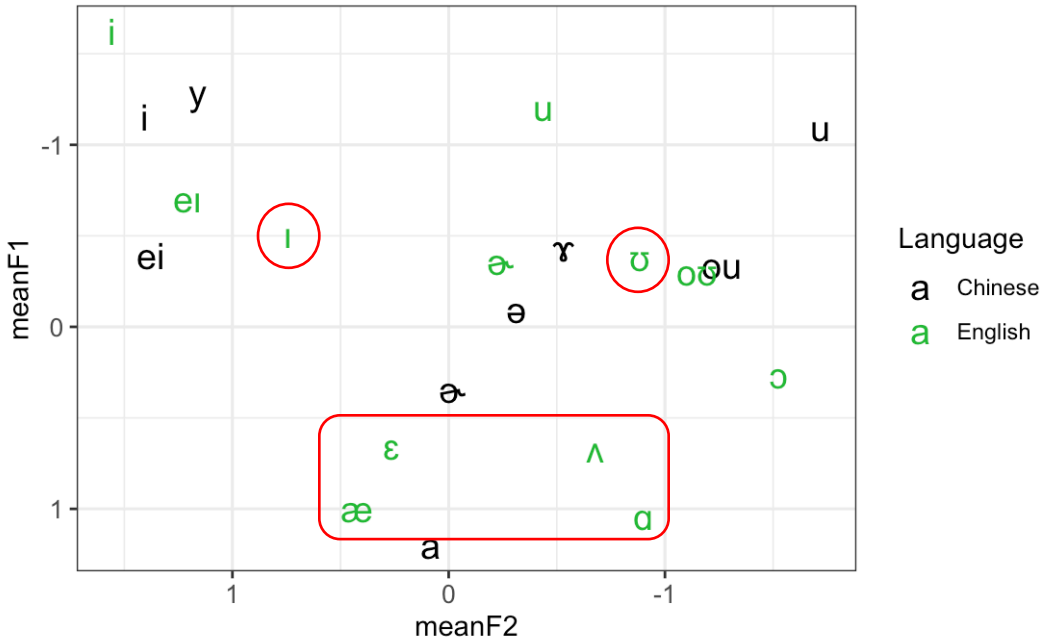


Acoustic Vowel Spaces: Monophthongs

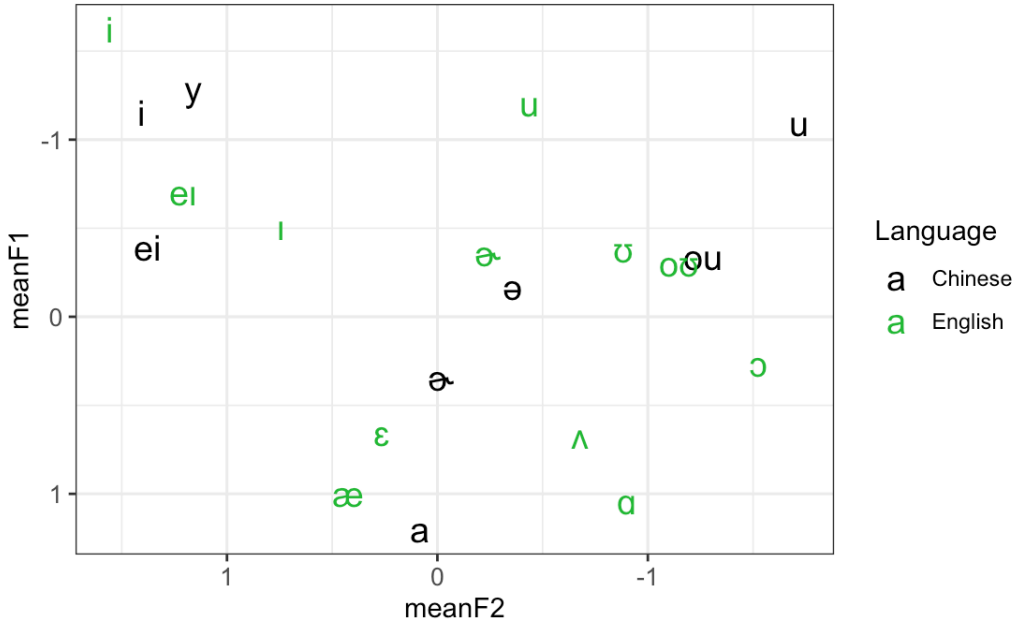
Mandarin Pinyin vs. English Phonemes



Mandarin Phoneme(10) vs. English Phonemes



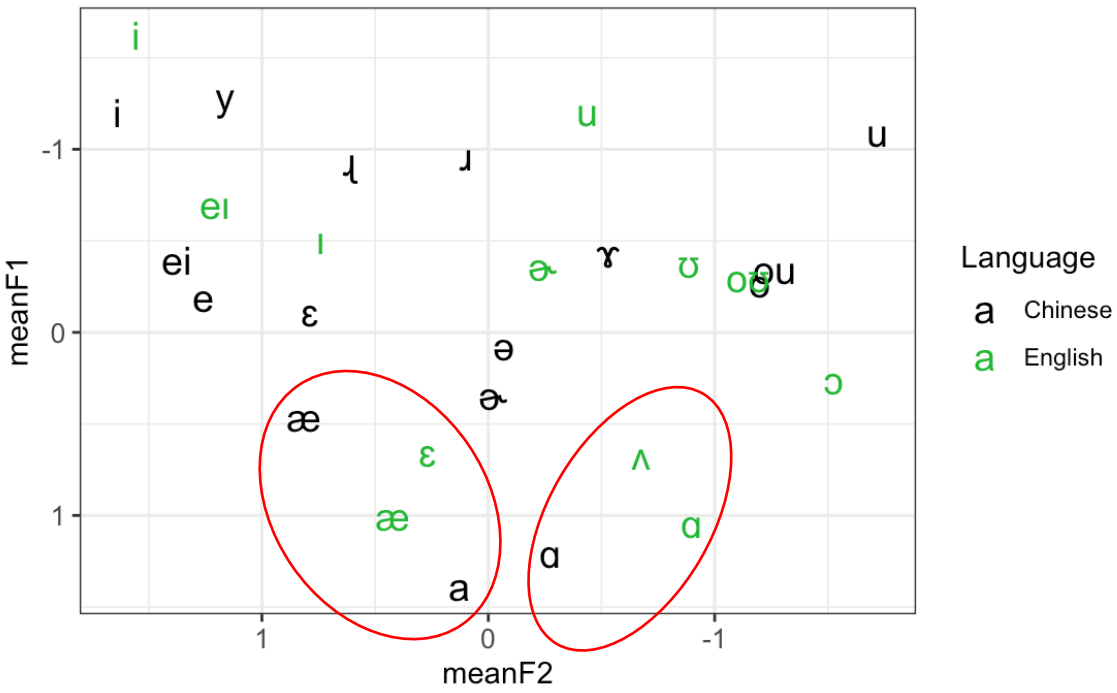
Mandarin Phoneme(11) vs. English Phonemes



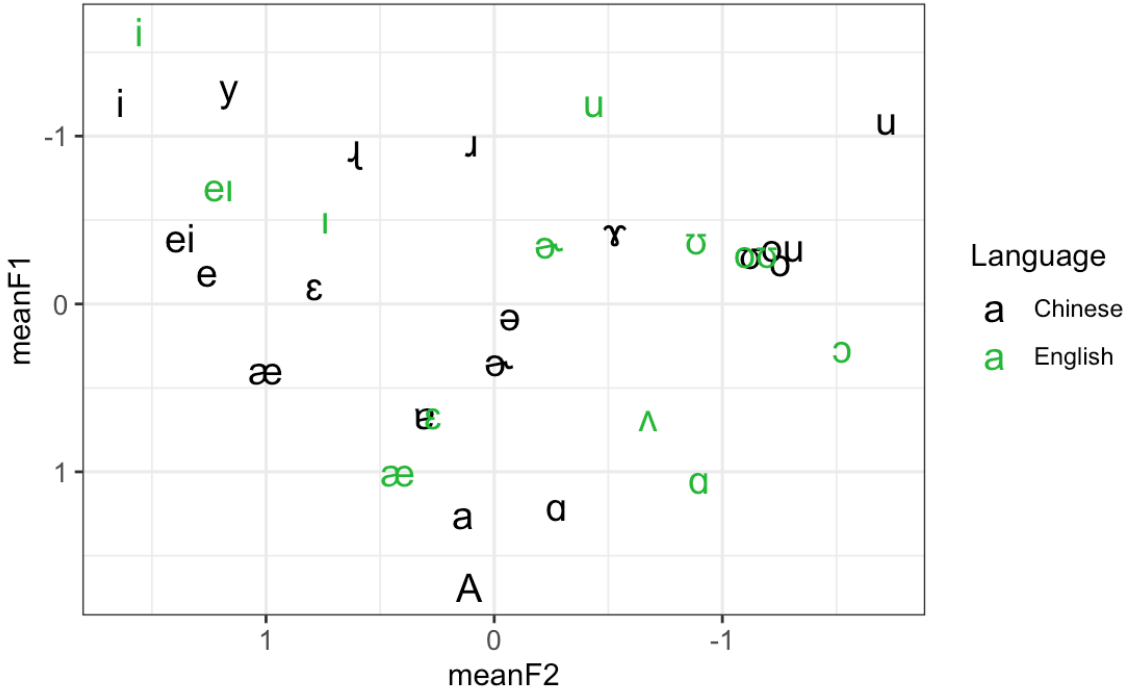
Acoustic Vowel Spaces: Monophthongs

- Same IPAs but different acoustic qualities across languages

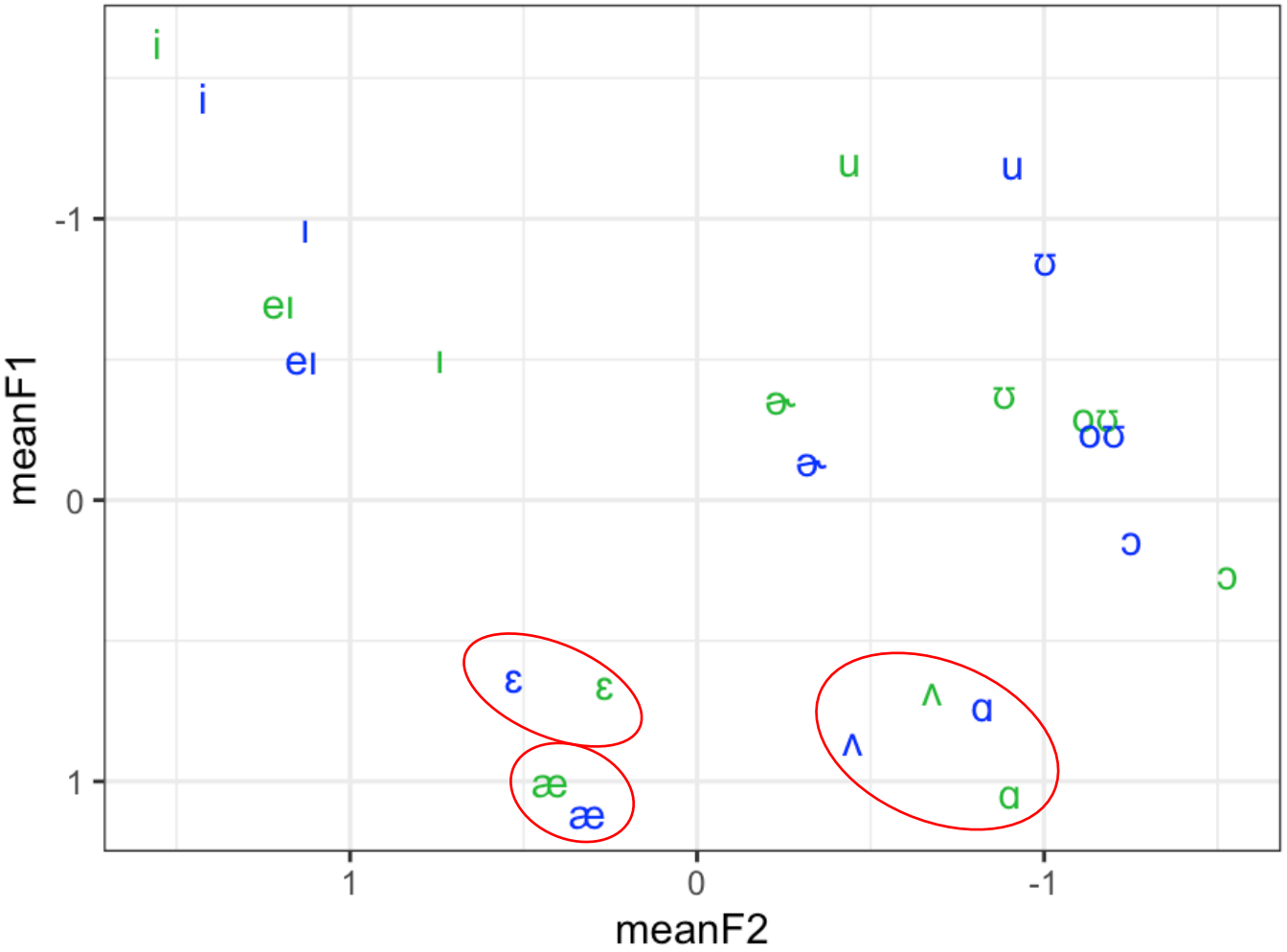
Mandarin Allophone(18) vs. English Phonemes



Mandarin Allophone(22) vs. English Phonemes



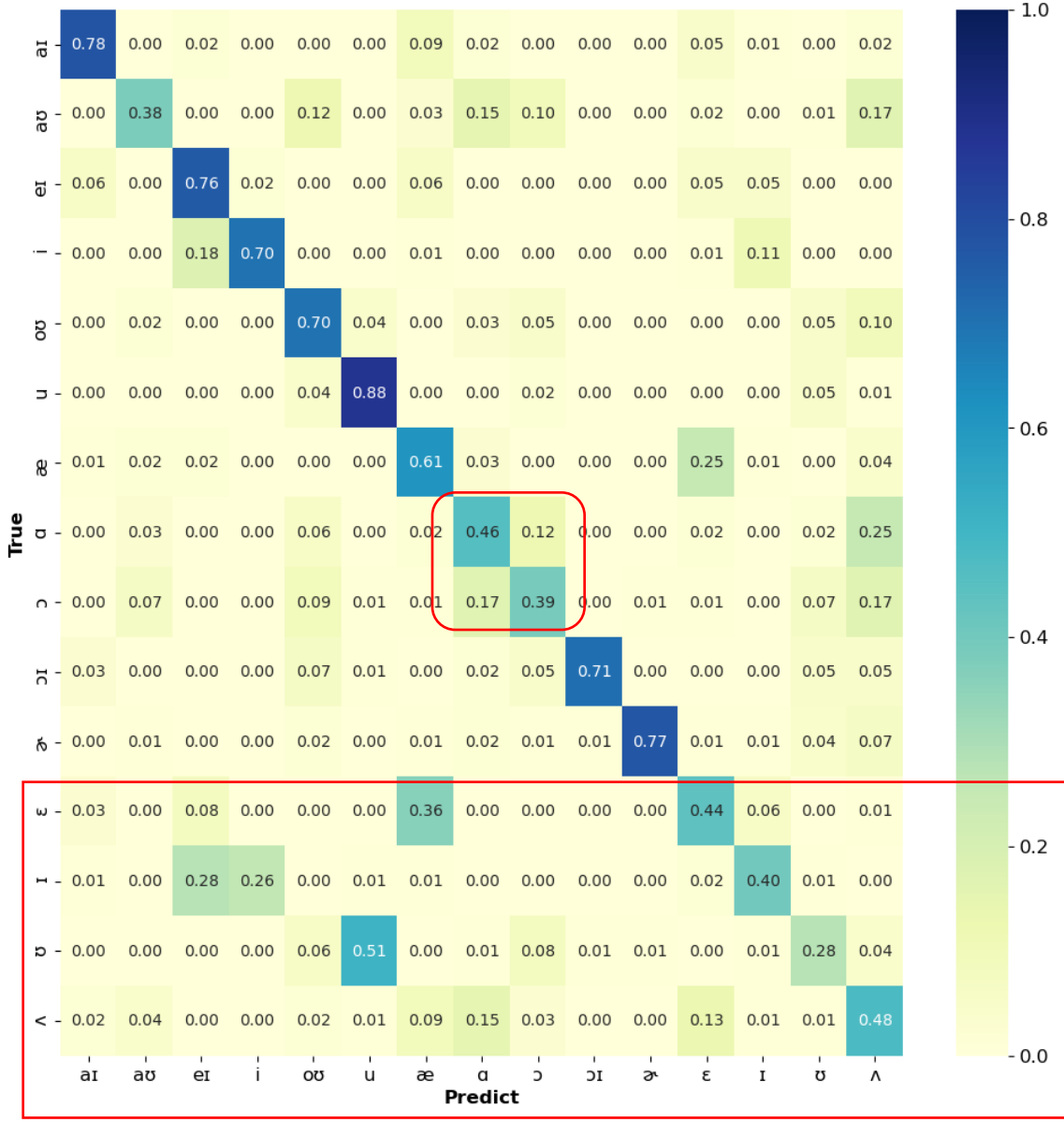
Native English vs. L2 English: Monophthongs



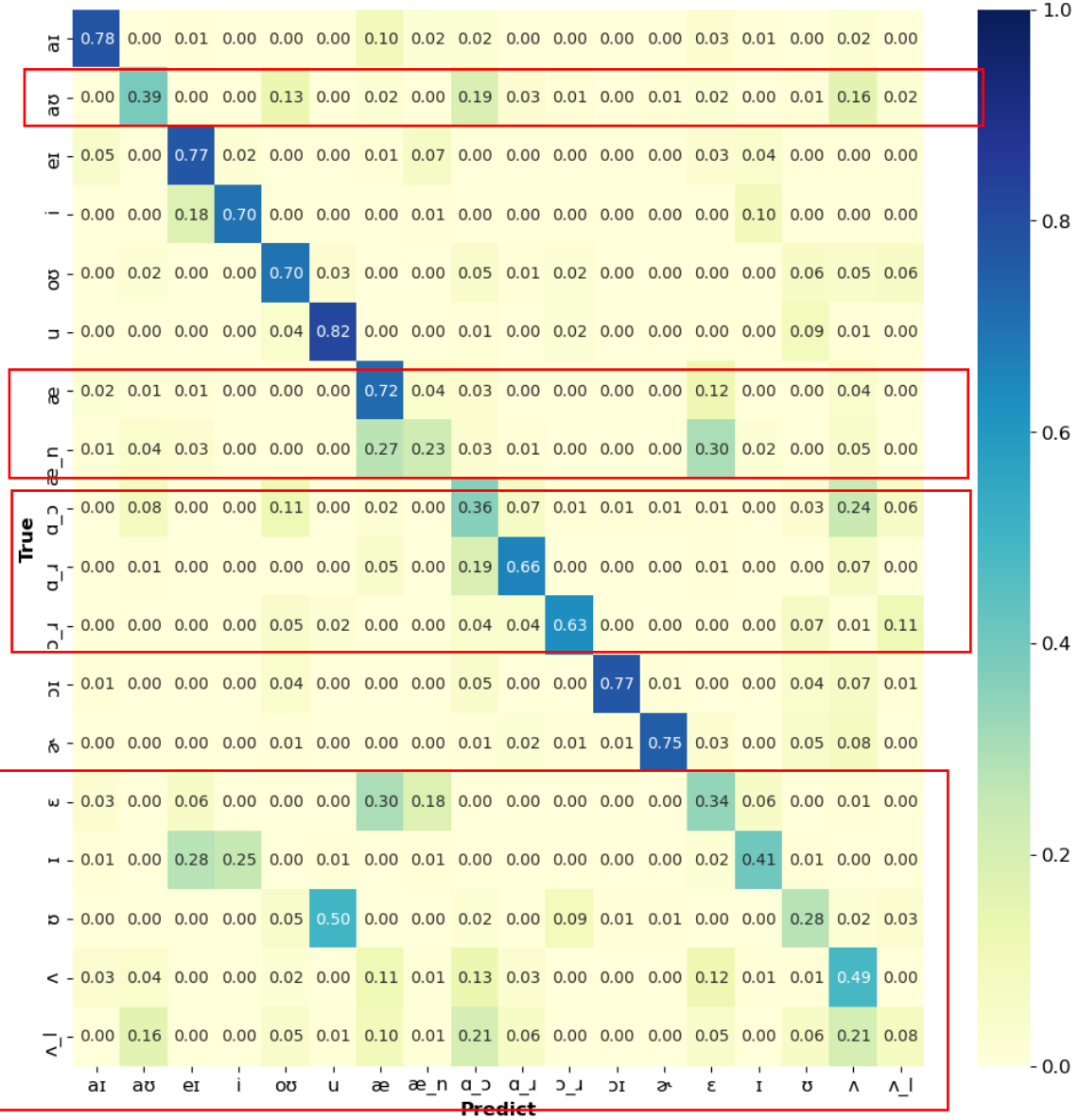
- /ɛ/ is fronter, /æ/ is lower and more back, /ɑ/ is higher and /ɪ/ is fronter and higher : suggesting assimilation effect on both phonemic and allophonic levels

L2E Classification Results

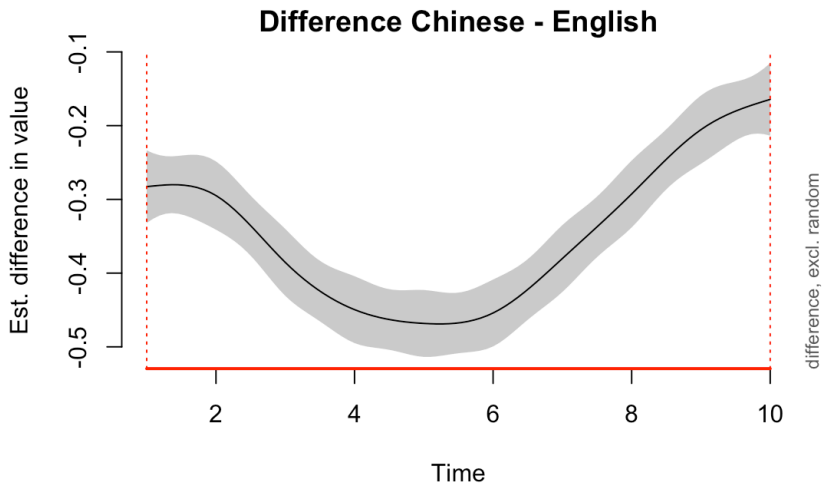
Phoneme: 61.1%

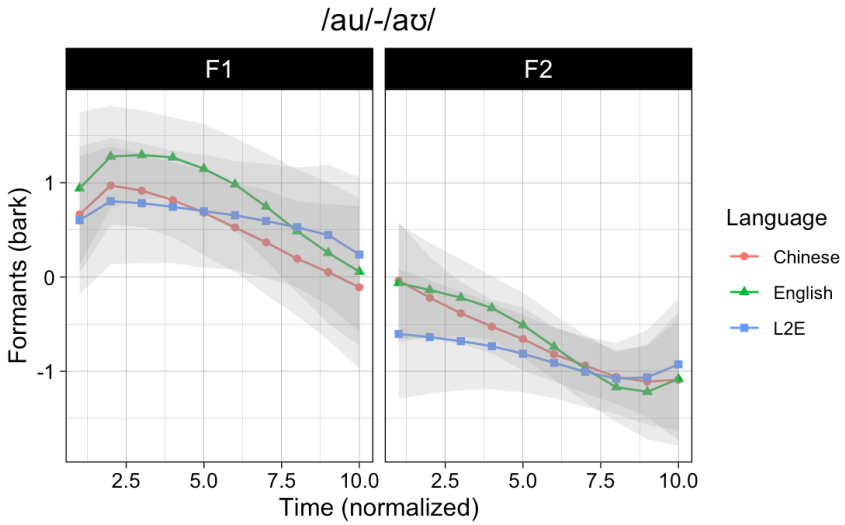


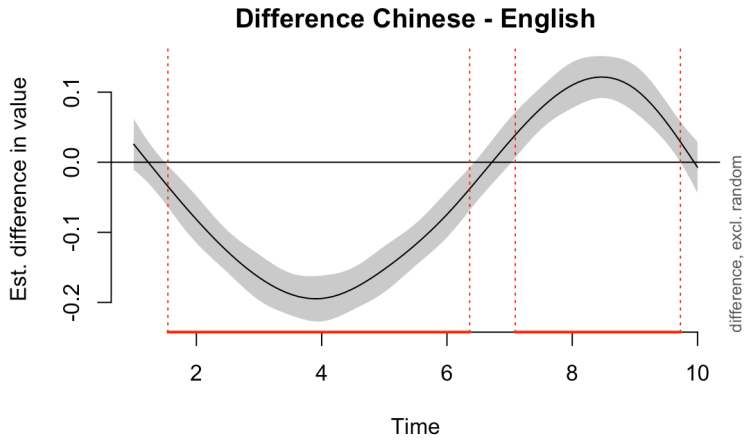
Allophone: 60.1%

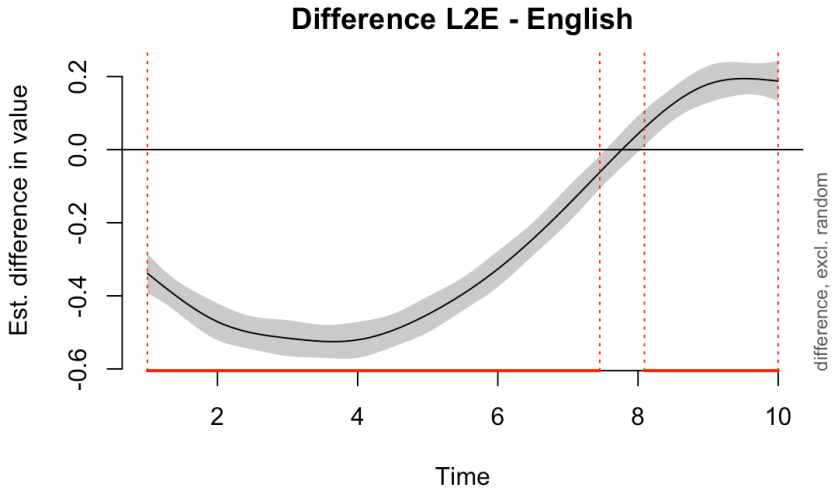


Diphthongs: /au/-/aʊ/

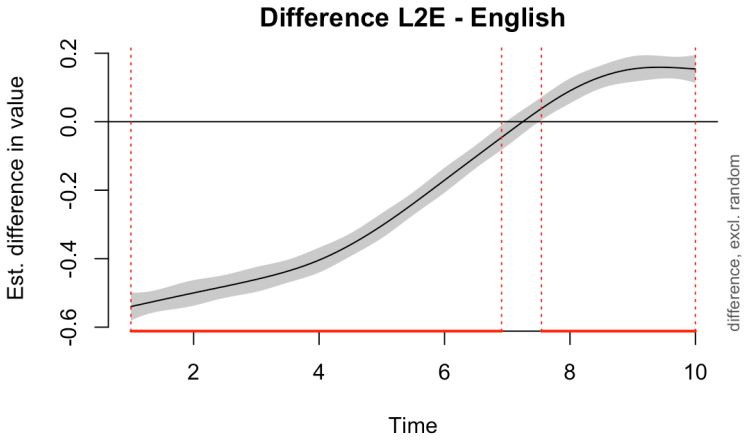






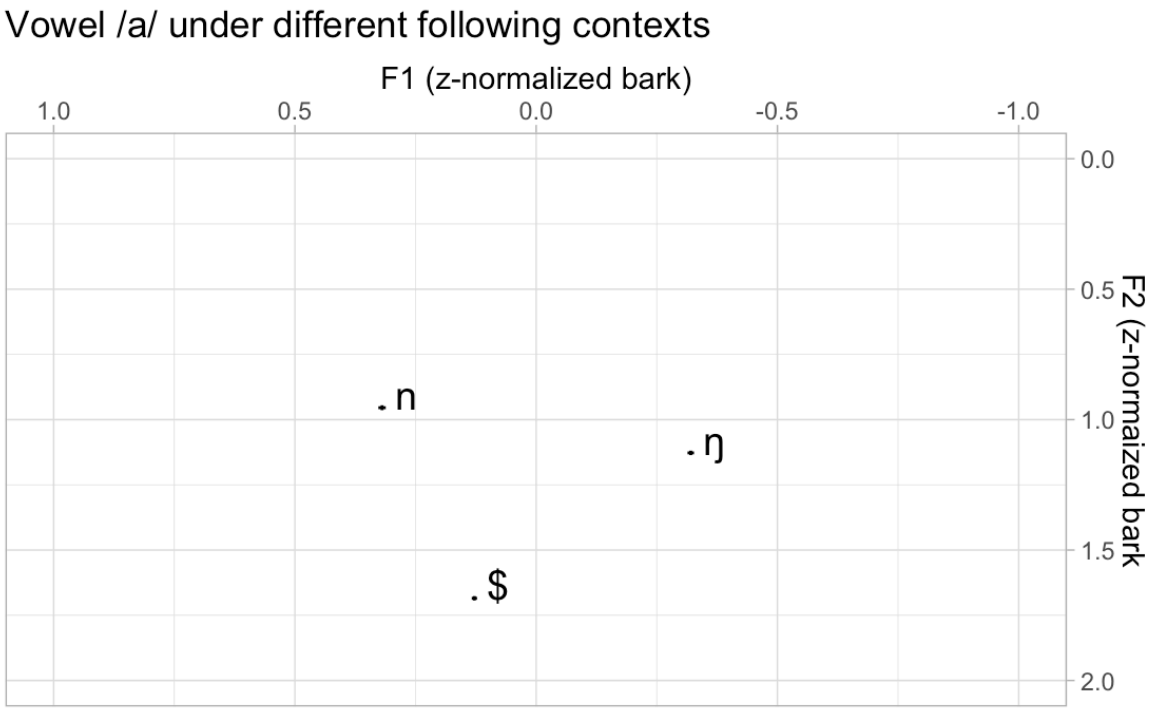


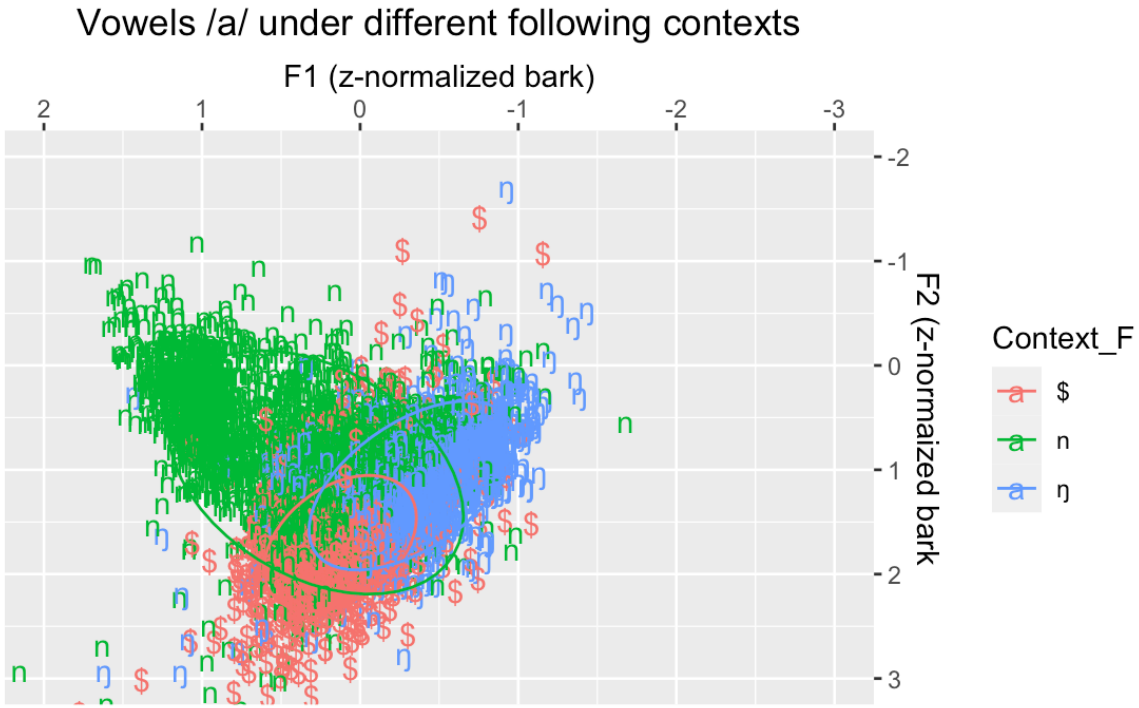
- Chinese vs. English
  - Lower F1 for Chinese /au/
  - Lower F2 for Chinese /a/ and higher F2 for Chinese /u/
- L2E vs. English
  - Lower F1 for L2E /a/ and higher F1 for L2E /ʊ/
  - Lower F2 for L2E /a/ and higher F2 for L2E /ʊ/



Co-articulation and L2 acquisition

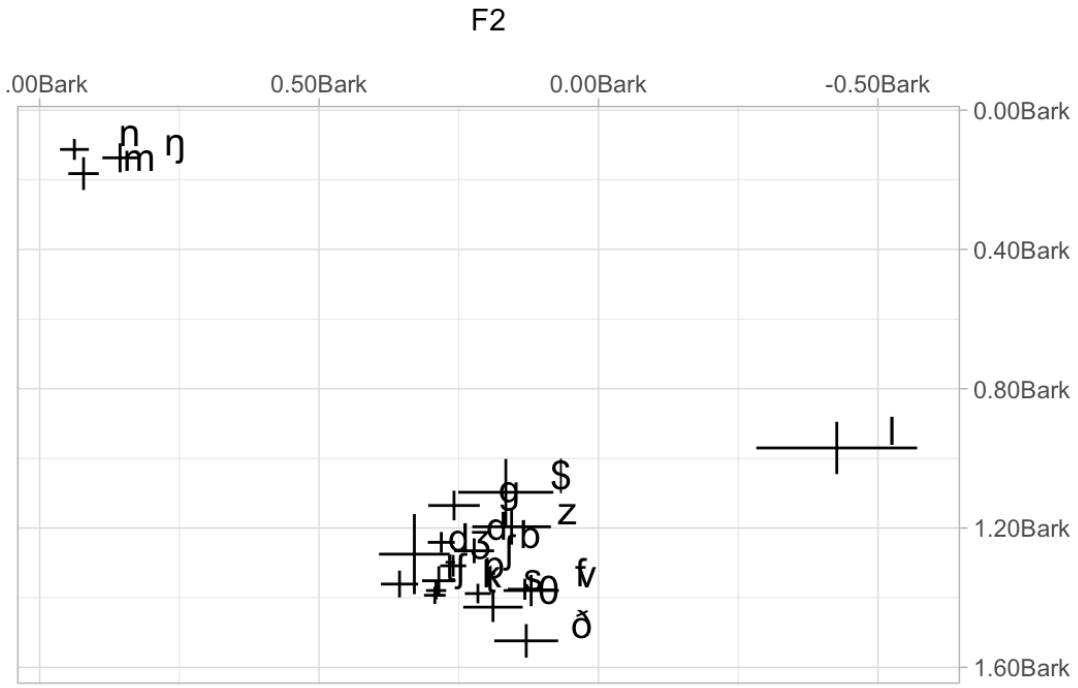
- Mandarin /a/ under nasal contexts
  - /a/ is raised under both nasal contexts (lower F1 values)
  - /a/ is more back when followed by the velar nasal /ŋ/

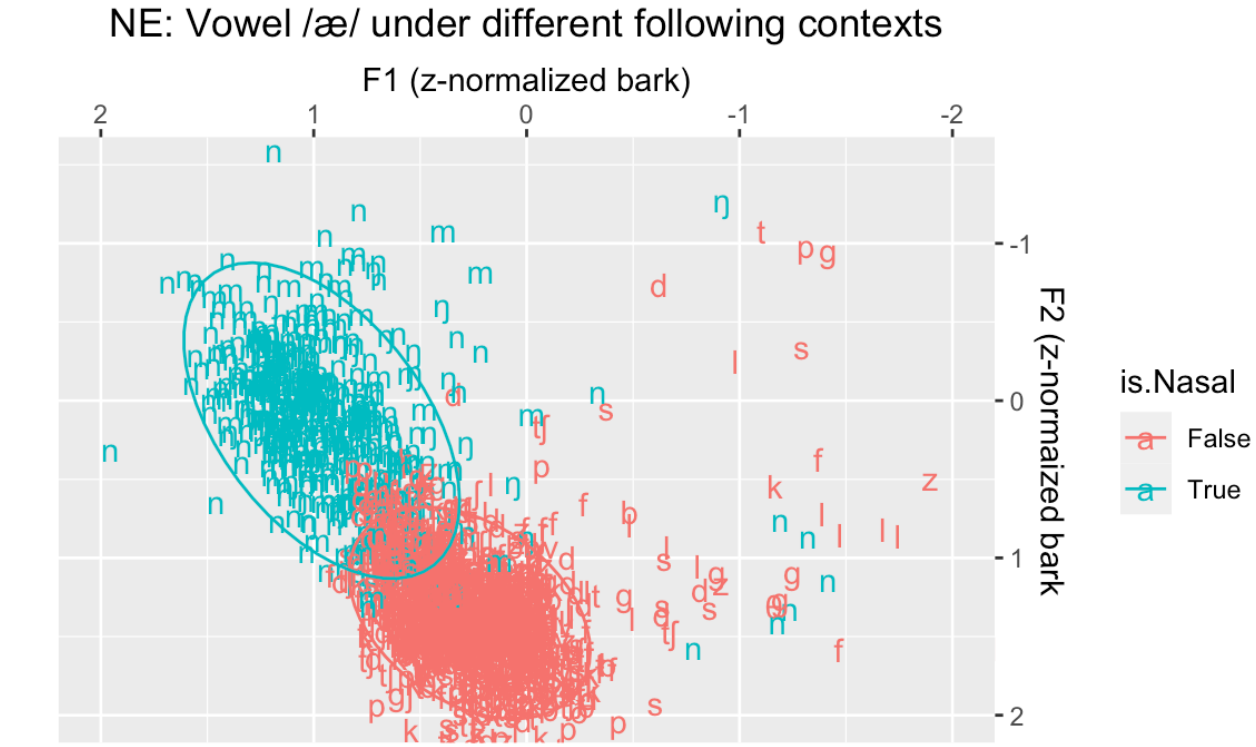




Co-articulation and L2 acquisition

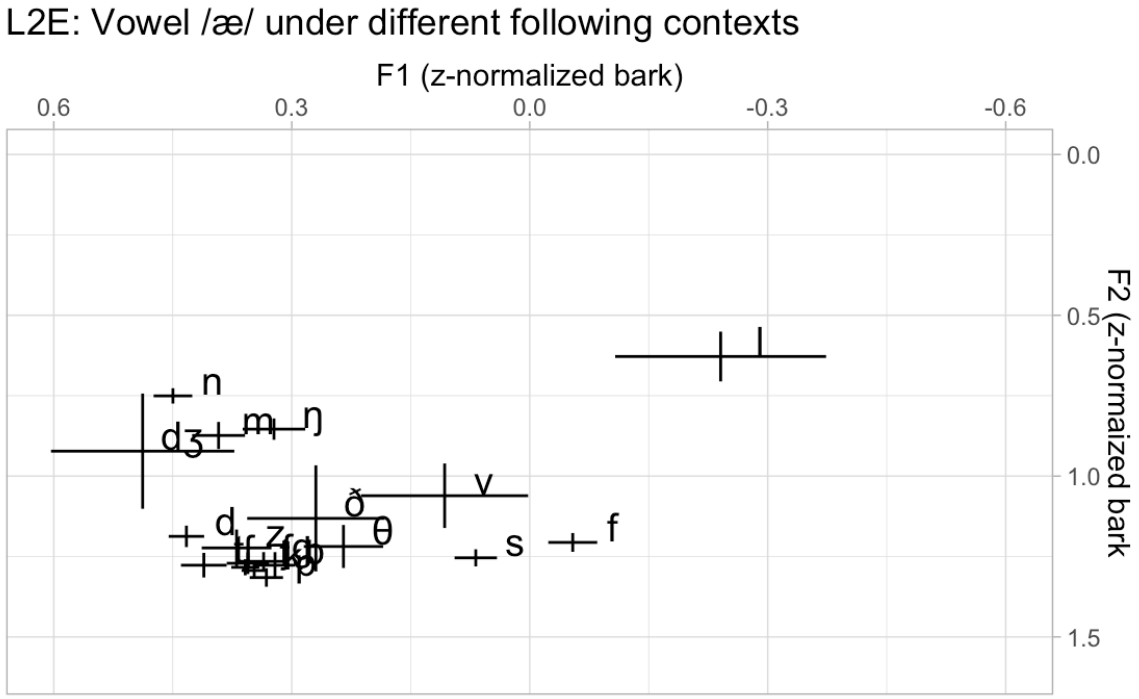
- Native English /æ/
  - Three different clusters, æ_nasal, æ_l, æ_other
  - /æ/ is raised and fronted for all nasal conditions (lower F1 values and higher F2 values)
  - There is a slight difference between different nasal conditions.

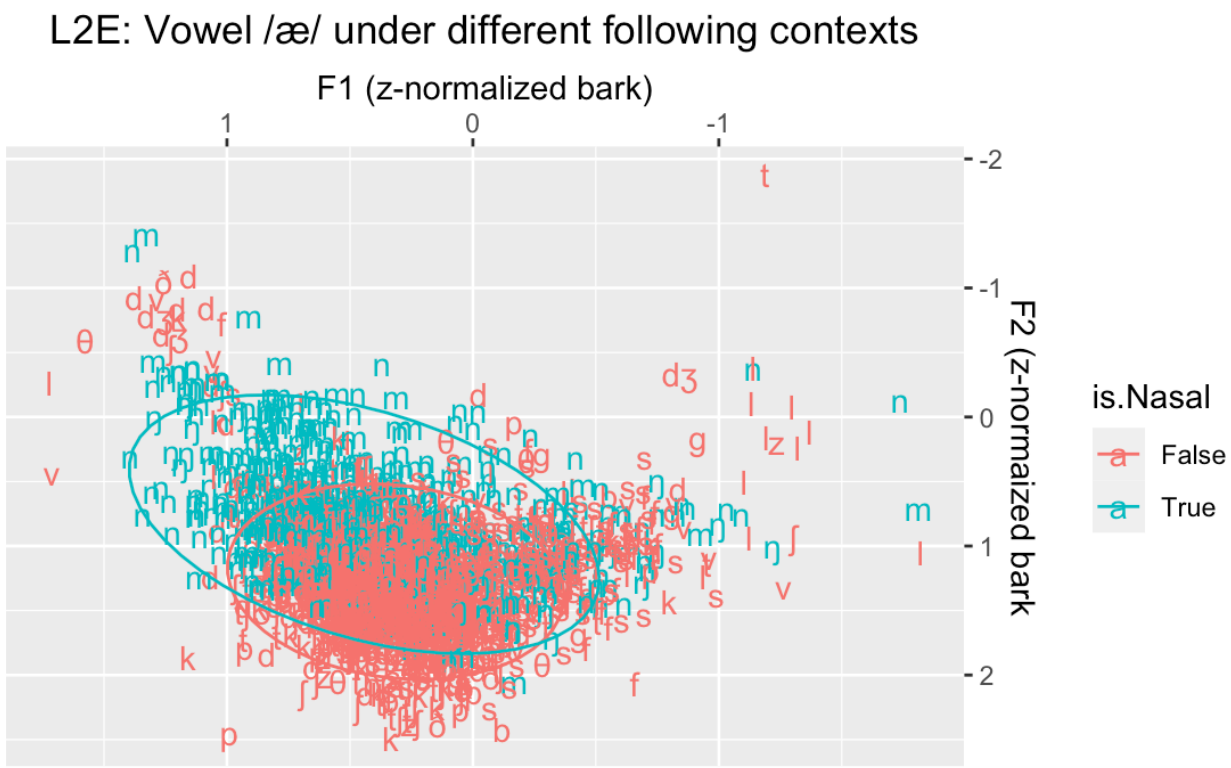




Co-articulation and L2 acquisition

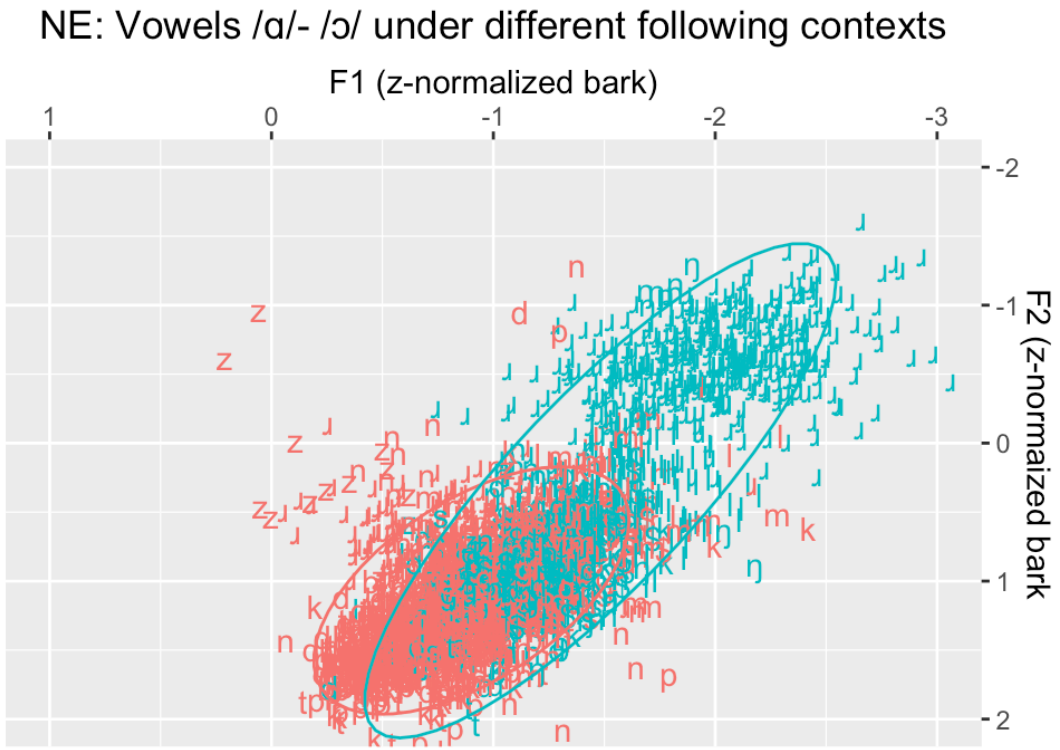
- L2E /æ/
  - There are approximately three different clusters, æ_nasal, æ_l, æ_other
    - The difference between æ_nasal, and æ_other is smaller but the difference among æ_nasal is bigger.

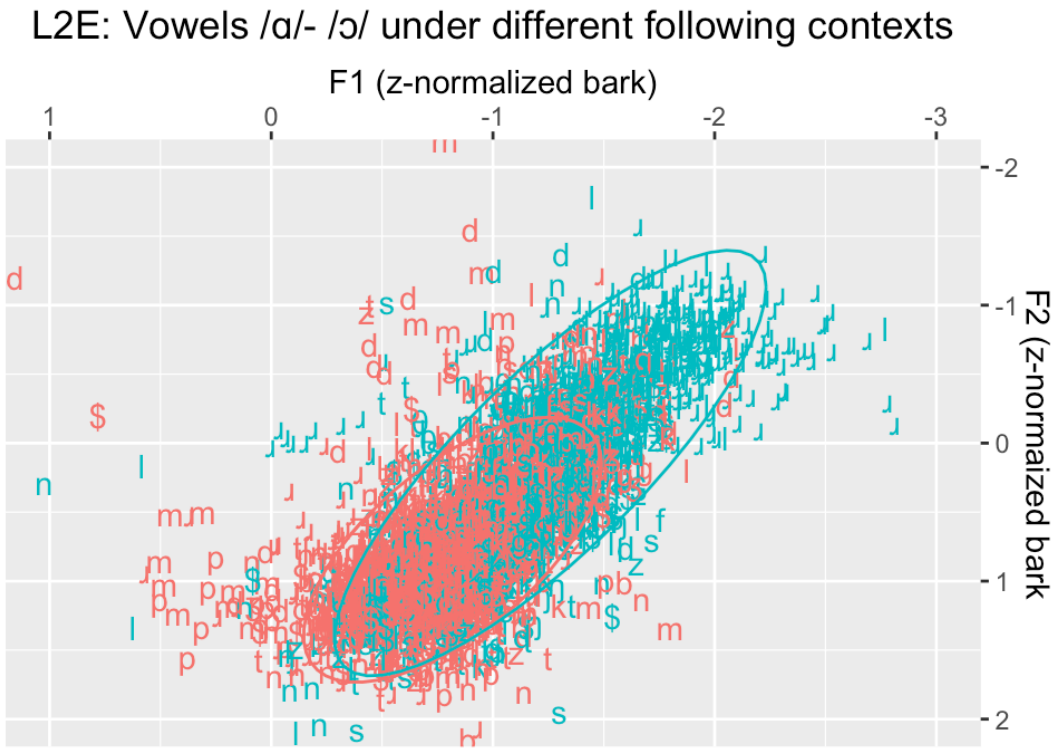




Co-articulation and L2 acquisition

- L2E: /ɑ/-/ɔ/
  - There is partial /ɑ/-/ɔ/ merger in both cases.
  - The difference between ɑ_ɹ vs. ɑ-ɔ, and that between ɔ_ɹ v.s ɑ-ɔ is smaller in native English than in L2 English.





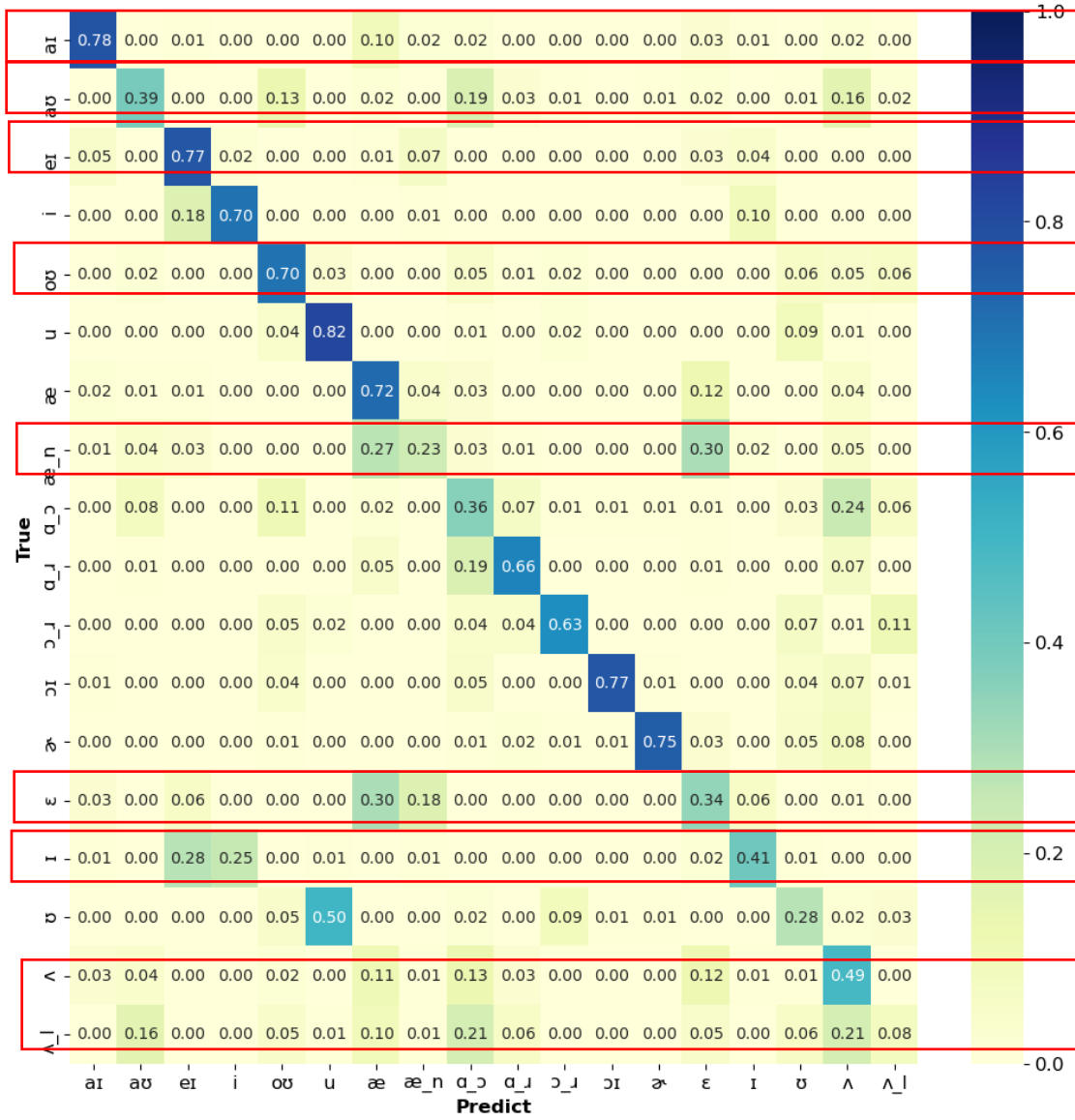
Assimilation

- English to Chinese
  - How English phones are assimilated to Chinese phones under each assumption
    - PCA transformation
    - Vowel inventory levels

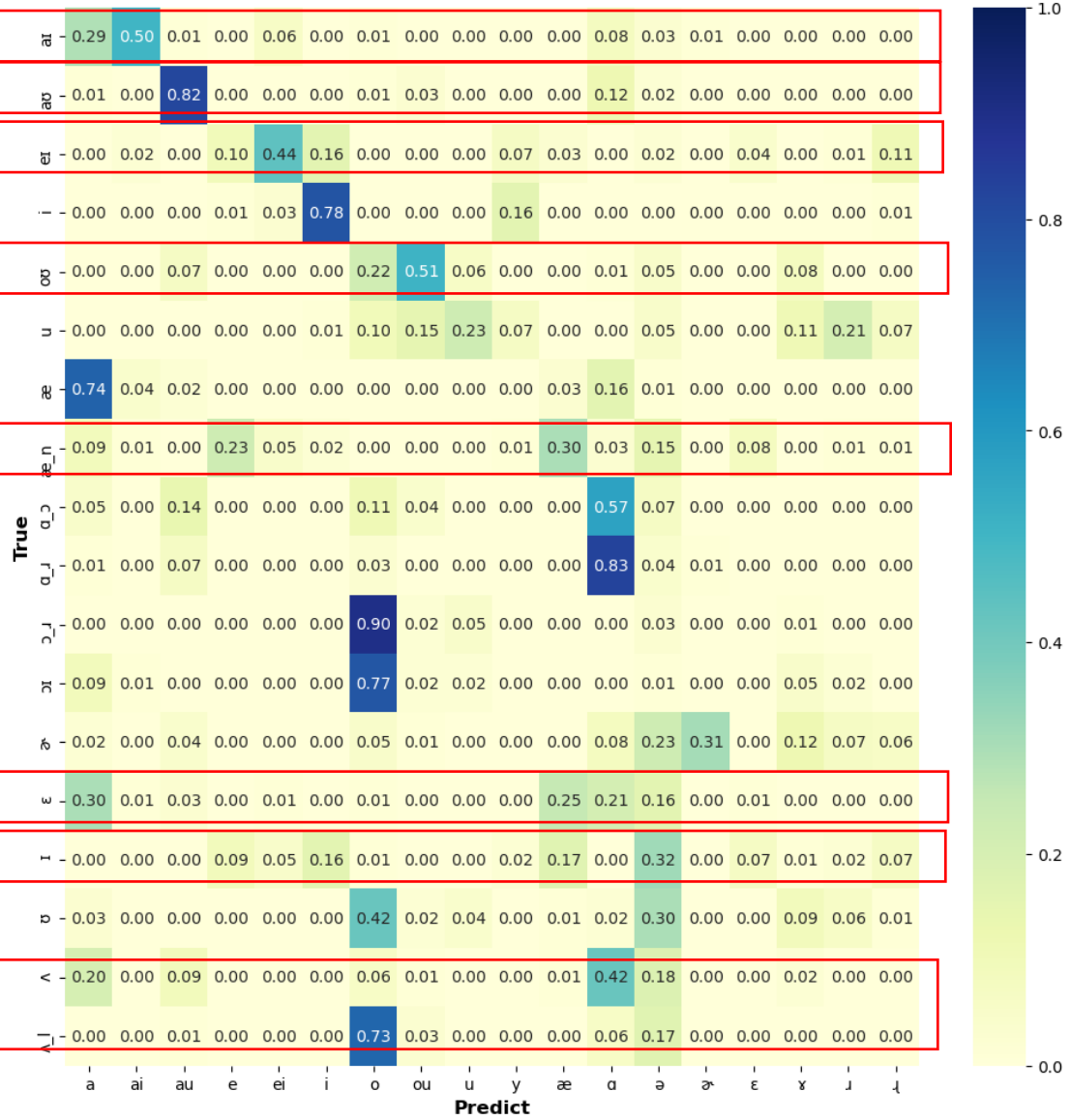
- English to English
  - How transformed English phones are assimilated to English phones under each assumption
    - PCA transformation
    - Vowel inventory levels

Assimilation Results: English to Chinese

L2E classification results at allophone level

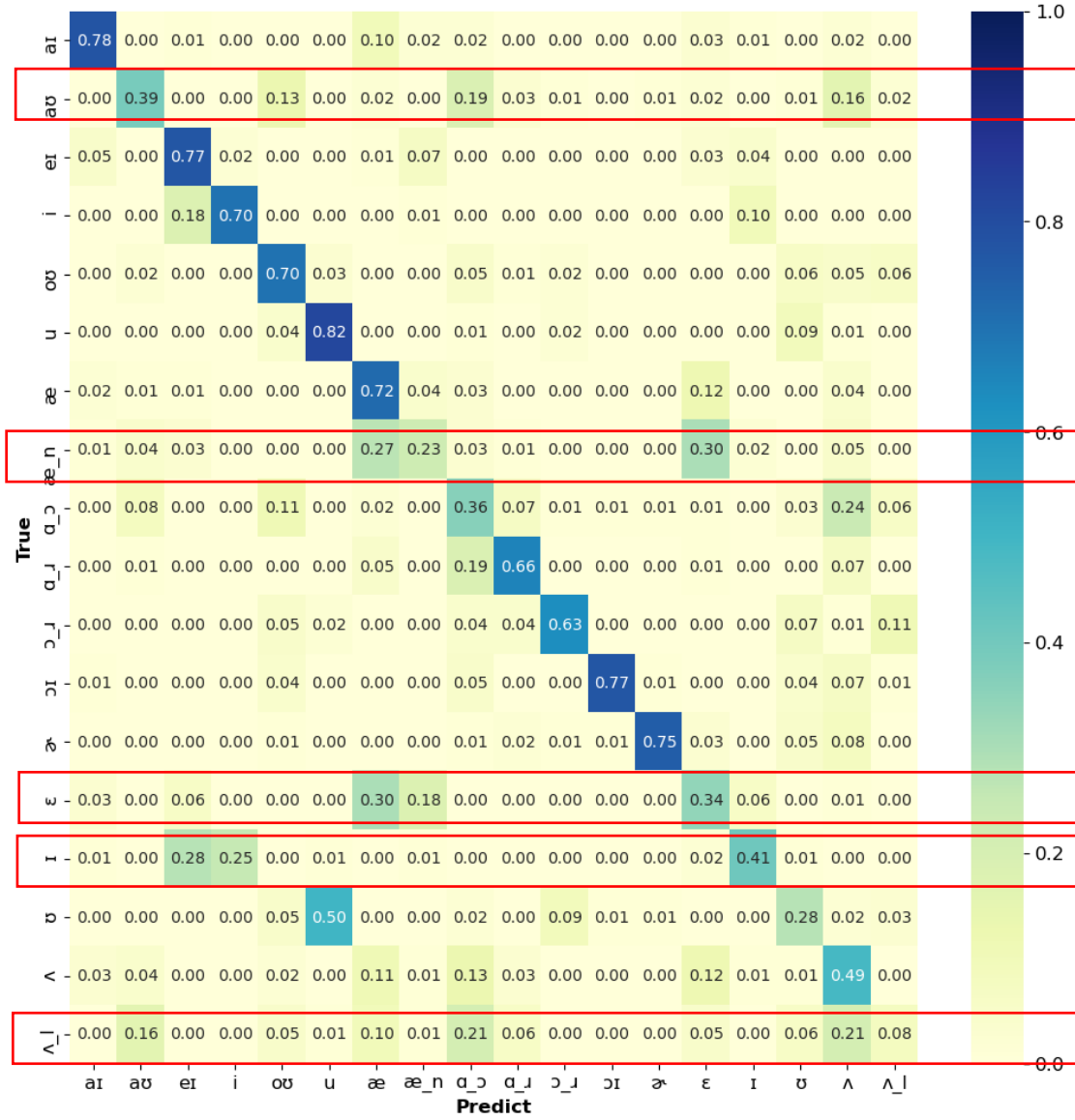


Assimilation results using PCA3 and Allophone(18)

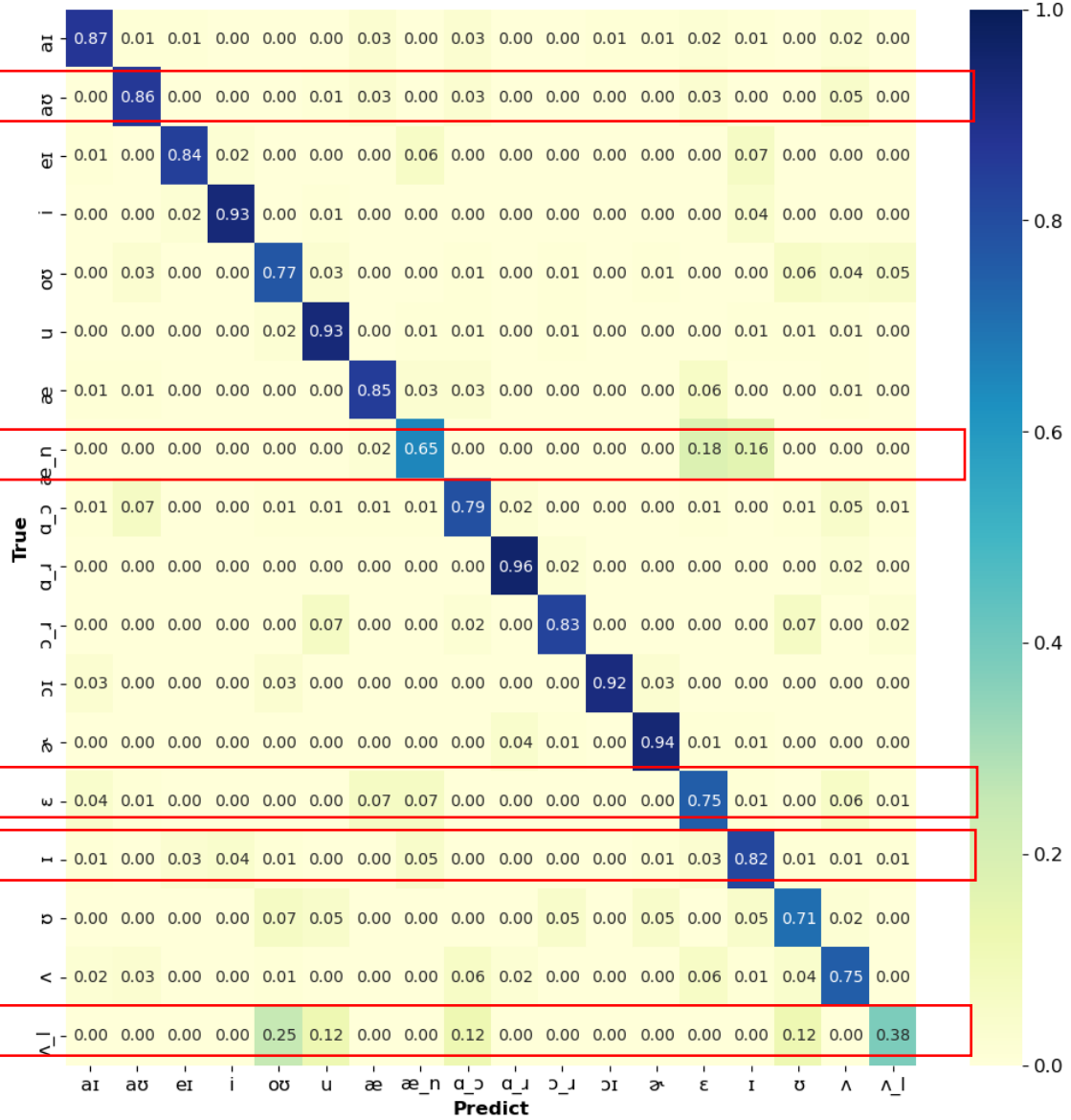


Assimilation Results: English to English

L2E classification results at allophone level



Assimilation results using PCA3 and Allophone(18)



Conclusion and Future work

- Conclusion
  - L1 and L2 are likely to share a common phonological / phonetic space
  - Assimilation could happen at both phonemic and allophonic levels
  - Our approach is effective in simulating L2 assimilation and in automatic prediction of L2 pronunciation errors
  - The English-to-Chinese assimilation approach can account for more pronunciation errors than the English-to-English assimilation approach so far.
- Future work
  - Analysis of L2 error patterns in more detail
  - Analysis by different L2 proficiency levels
  - Quantitative assessment of assimilation results
  - Pre-processing of pronunciation errors introduced by orthography

Thank you!
Any questions?