

Automatic Pronunciation Assessment of Non-native English based on Phonological Analysis

C. D. Rios-Urrego¹, D. Escobar-Grisales¹, S. A. Moreno-Acevedo¹, P. A. Perez-Toro^{1,2}, E. Nöth², and J.R. Orozco-Arroyave^{1,2}

¹GITA research Lab, Faculty of Engineering, University of Antioquia, Medellín, Colombia
²Pattern Recognition Lab, Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany



Introduction

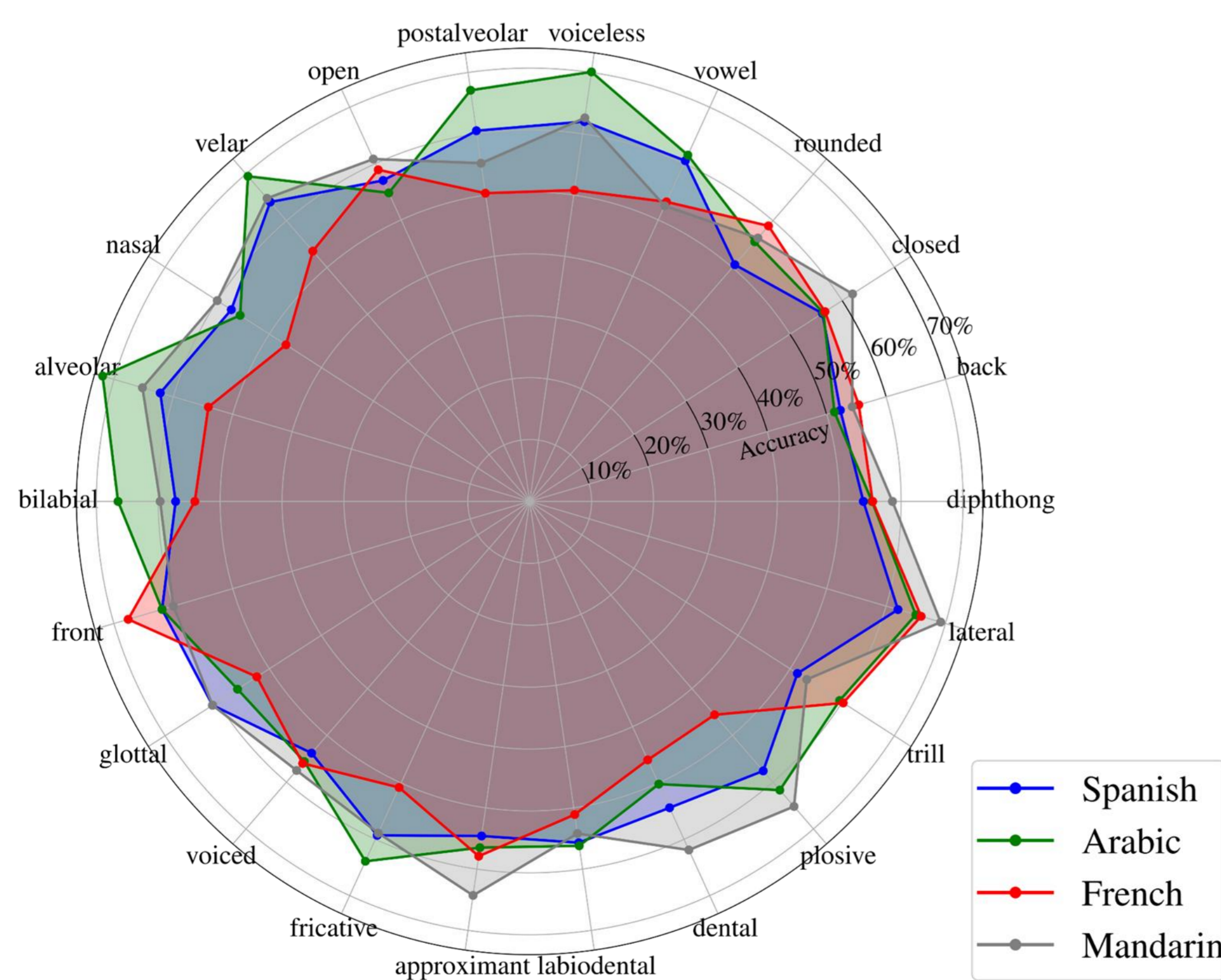
- English is the official language in over 50 countries and is considered the language of international communication
- Computer-based assessments analyze speech to objectively evaluate English levels, providing feedback for users to identify strengths and weaknesses
- An approach to the classification and evaluation of phonological classes for non-native speakers is proposed in this work
- Two different scenarios were performed:
 - Classification between native English speakers vs. non-native English speakers
 - Phoneme-by-phoneme pronunciation level feedback generation by phonological class

Database

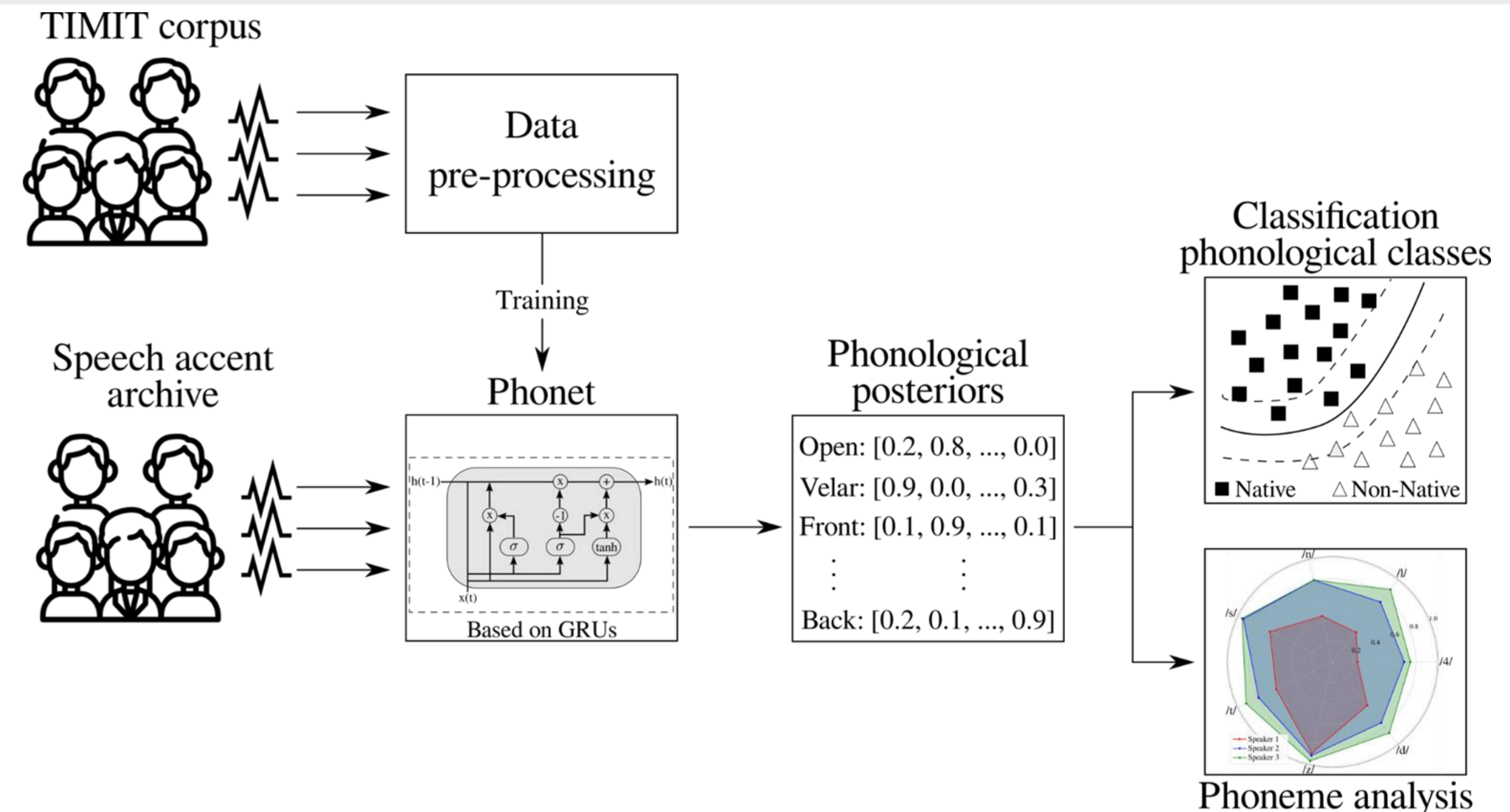
- TIMIT Corpus:
 - 2342 sentences read by 630 speakers with different dialects of American English
 - This corpus includes time-aligned orthographic, phonetic, and word transcriptions
 - It was used to train and evaluate Phonet
- Speech Accent Archive:
 - 2140 speech samples, each from a different talker reading the same reading passage in English
 - We considered five groups of participants: English, Spanish, Arabic, Mandarin, and French
 - Each non-native language was paired with a native group based on statistical tests on age and gender

Classification of Phonological Classes

Native language	Phonological classes	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-score (%)
Arabic	All	80.6 ± 2.5	85.9 ± 2.7	75.3 ± 2.9	80.5 ± 2.5
	Alveolar	71.9 ± 3.1	68.4 ± 1.0	75.3 ± 5.7	71.8 ± 3.1
Mandarin	All	74.2 ± 2.8	78.5 ± 4.5	69.9 ± 1.9	74.1 ± 2.8
	Lateral	69.2 ± 3.2	71.7 ± 3.0	66.8 ± 3.5	69.2 ± 3.2
Spanish	All	72.0 ± 1.8	71.9 ± 2.0	72.1 ± 2.9	72.0 ± 1.8
	Velar	64.0 ± 0.9	73.6 ± 2.9	54.4 ± 2.4	63.7 ± 0.9
French	All	67.6 ± 2.0	74.3 ± 3.4	61.0 ± 5.0	67.4 ± 2.1
	Front	67.6 ± 1.5	79.7 ± 5.4	55.6 ± 5.0	67.1 ± 1.5



Methodology

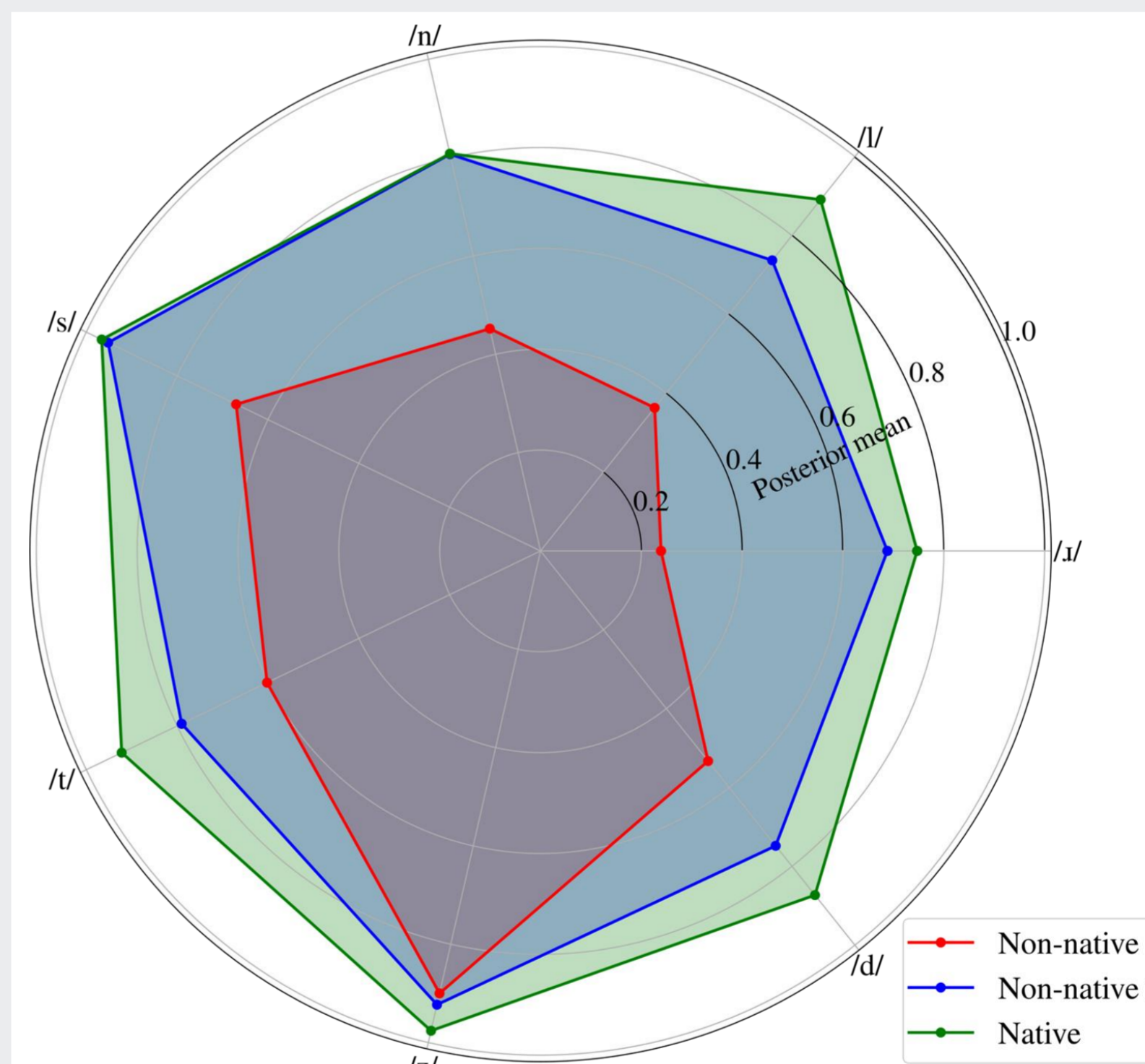


We trained Phonet using TIMIT corpus audios and transcriptions. Then, we processed recordings from native and non-native speakers in the Speech Accent Archive, calculating phonological posteriors for each class. Our approach involved: (i) classifying non-native speakers vs. native speakers sets, and (ii) conducting a phoneme-level analysis to provide feedback on individual speaker pronunciation

Phonological analysis

- Phonological features are used to model the information about the place and manner of articulation of a speaker
- The phonological posteriors are the posterior probability of a speech frame belonging to one or more phonological classes
- Phonet is a toolkit to estimate the phoneme articulation precision of different speakers and used these posteriors to classify native and non-native English speakers
- 22 phonological classes were considered in this study (Phonet's accuracy was 92.5 ± 3.2%)

Phoneme Analysis



- Radar plot shows the mean posterior for three speakers for the Alveolar class
- The speaker receives feedback about his/her pronunciation performance

Conclusions

- Results indicate the potential to differentiate non-native languages from native English, achieving accuracies of 67.6% to 80.6%
- It was possible to identify the most discriminant phonological classes are alveolar, lateral, velar, and front
- Future work will consider training a multilingual system that allows the automatic evaluation of pronunciation in different languages

Contact

Cristian David Rios-Urrego, M.Sc.
 GITA Lab, Faculty of Engineering,
 University of Antioquia, Medellín, Colombia
 cdavid.rios@udea.edu.co

Acknowledgement

This work received funding from UdeA grant # ES92210001 and CODI grant No. PI2023-58010, and PRG2017-15530