Measuring the Impact of Segmental Deviation on Perceptions of Accentedness using Gradient Phonological Class Features

Nitin Venkateswaran, Rachel Meyer, and Ratree Wayland
Department of Linguistics, University of Florida
{venkateswaran.n, rmeyer2, ratree@ufl.edu}

BACKGROUND & OBJECTIVES

- **Hindi/Indian English**: The retroflex stop [ʈ], labiovelar approximant [ʋ], and rhotic tap [ɾ] are uniformly produced in Indian English (IE) varieties [1], including the English of L1 Hindi speakers (Hindi English)
- **Research Questions:** Can Hindi English (HE) speakers' segment deviations, perceived by native speakers of American English (AE), be measured using explainable phonological feature-based representations derived from deep neural networks?
- **Methods & Approaches:**
  - Train Phonet [2] on baseline IE&AE data to estimate phonological class probabilities of target HE segments.
  - 2-way ANOVAs investigating differences in phonological class probabilities between expected AE and realized HE segments as drivers of accent perception.
  - Using the **perceptual space** of the Phonet model, investigate associations between accent ratings and vector-based distances from target HE to baseline IE&AE segments via multinomial logistic regressions.

PHONET

- Bi-directional GRU model uses MFCC transformations of acoustic signals.
- Generates vector of phonological class probabilities as speech representations of phone segments.

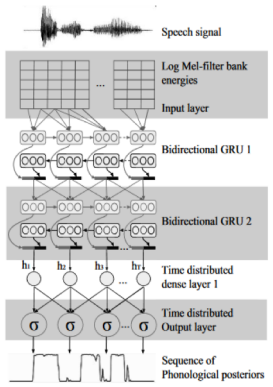

Figure 1. Phonet architecture [2]

CSLU FAE ACCENTED DATASET

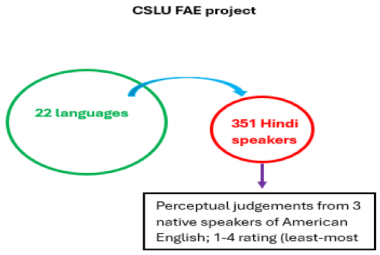

Figure 2. The CSLU FAE project & dataset [3]

PHONET TRAINING PROCEDURE

- ~270h IE&AE baseline ASR data [4][5][6][7]; 80:20 train-test split; 30 epochs with early stopping; multi-task setup with classification heads for each phonological class.
- Custom MFA [8] IE&AE acoustic models generate TextGrids with phone-level alignments for phonological class annotations.
- Averages frame-level phonological class probability vectors to derive phone-level representation.

RESULTS

HE [ɾ] vs. AE [ɹ]

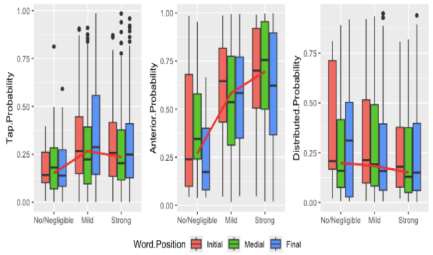

Figure 3: Phonological class probability distributions by accent rating and word position for the [tap], [anterior] and [distributed] classes contrasting realized HE [ɾ] vs. expected AE [ɹ] segments. 2-way ANOVAs show significant main effects of accent ratings on the probabilities of all three features.

RESULTS cont.

- **2-way ANOVA**: Main effects of accent rating on [anterior], [tap], and [distributed] probabilities
- **Logistic Regression**: Main effects of IE&AE distance; odds of *Mild* & *Strong* accents increases (decreases) with increasing distance from AE (IE) baselines

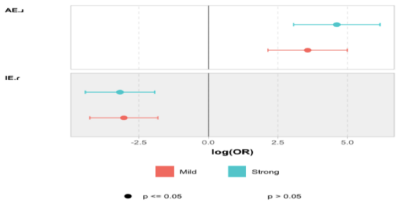

Figure 4: Multinomial logistic regression of HE [ɾ] distances from IE&AE baselines on accent ratings, with No/Negligible accent as reference level.

HE [ʋ] vs. AE [w]

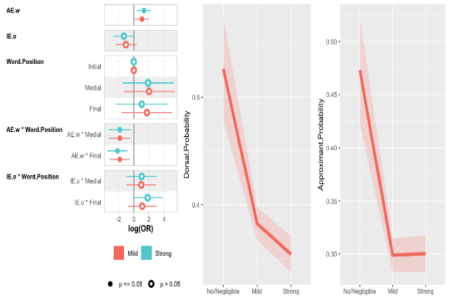

Figure 5: (Left) Multinomial logistics regression of HE [ʋ] distances from IE&AE baselines on accent ratings, with No/Negligible accent as reference level. (Center, Right) Word-initial interaction plots of [dorsal] and [approximant] probabilities by accent rating.

- **2-way ANOVA**: Interaction effects of accent rating and word position on [dorsal] & [approximant]: probabilities decrease word-initially as accent strength increases
- **Logistic Regression**: Interaction effects of AE distance with word position; odds of *Mild* and *Strong* accents higher word initially and medially with unit increase in AE distance.

RESULTS, cont.

HE [ʈ] vs. AE [t]

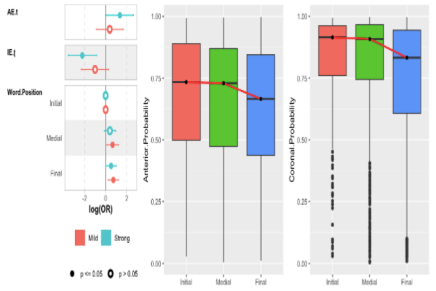

Figure 6: (Left) Multinomial logistics regression of HE [ʈ] distances from IE&AE baselines on accent ratings, with No/Negligible accent as reference level. (Center, Right) Probability distributions of [anterior]&[coronal] classes by word position

- **2-way ANOVA**: Main effects of word position on [anterior] and [coronal] probabilities: lower probabilities word-finally
- No main not interaction effects of accent ratings on probabilities; supports research showing at-chance discrimination of retroflex vs. dental Hindi stops by AE listeners [9,10]
- **Logistic Regression**: Main effects of IE&AE distance; odds of *Mild* & *Strong* accents increases (decreases) with increasing distance from AE (IE) baselines, with prominent word-final effects
- Suggests other factors in retroflex stop segments contribute to accent perception

WORKS CITED

[1] Caroline Wiltshire. 2020. "Uniformity and Variability in the Indian English Accents", monograph for World English series
[2] Vásquez-Correa, J., Klumpp, P., Orozco-Arroyave, J. R., and Nöth, E. (2019). "Phonet: A tool based on gated recurrent neural networks to extract phonological posteriors from speech," in Proc. Interspeech 2019, Graz, Austria, pp. 549–553.
[3] Lander, T.CSLU: Foreign Accented English Release 1.2 LDC2007S08. Web Download. Philadelphia: Linguistic Data Consortium 2007.
[4] V. Panayotov, G. Chen, D. Povey and S. Khudanpur, "Librispeech: An ASR corpus based on public domain audio books," 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane, QLD, Australia, 2015, pp. 5206-5210
[5] G. Zhao, S. Sonsaat, A. Silpachai, I. Lucic, E. Chukharev-Hudilainen, J. Levis, and R. Gutierrez-Osuna. 2018. L2-ARCTIC: A Non-native English Speech Corpus. In Proc. Interspeech 2018, pages 2783–2787
[6] Common Voice: A Massively-Multilingual Speech Corpus arXiv:1912.06670 in Elements, Cambridge University Press
[7] Arun Baby, Anju Leela Thomas, N. L. Nishanthi, and TTS Consortium. 2016. Resources for Indian languages. In CBBLR Community-Based Building of Language Resources, pages 37–43, Brno, Czech Republic. Tribun EU
[8] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger. 2017. Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. In Proc. Interspeech 2017, pages 498–502
[9] Linda Polka. 1991. Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. The Journal of the Acoustical Society of America, 89(6):2961–2977
[10] John S. Pruitt, James J. Jenkins, and Winifred Strange. 2006. Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. The Journal of the Acoustical Society of America, 119(3):1684–1696