Impacts of native accent variant processing on human language technology: Integrating linguistic, psychological, and AI approaches.

Cristal Giorio, Jason Lucas, Janet G. Van Hell
The Pennsylvania State University

LinDiv is hosted at Penn State through NSF grant #2125865

Introduction

- A native language has different accent variants
- Accent variants within a language have distinct sound features [1]
  Ex. Mexican Spanish vs Puerto-Rican Spanish
- Studies tend to treat a given native accent as a collective accent of that language [2,3,4,5]

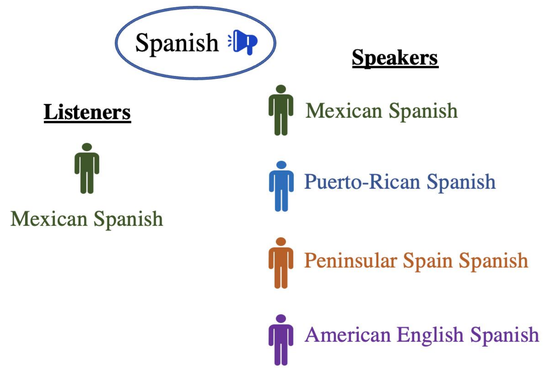
Topic 1: Are accent variants being processed differently by native listeners? Are there processing costs when listening to mismatched variants?

- Spoken language identification (LID) is a preprocessing step in language processing systems
- Language processing systems leveraged LID to identify the spoken language [6]
- LID progress in Automatic Speech Recognition (ASR) is largely due to emergence of :
  Large labeled multilingual speech corpora [6]
  Self-supervised pretrained models [7]
- ASR produces more errors for non-native speech [8, 9, 10]

Topic 2: What AI/NLP techniques can be explored to improve language ID accuracy for speech with strong regional or non-native accents?

Neurocognitive Approach

Lexico-Semantic Processing



★Listening to mismatched native speech can lead to processing costs

EEG

Findings

- Successful semantic access for matched native variant
- Unsuccessful semantic mapping for unfamiliar and less intelligible variant (Puerto Rican)

Applications to Human Language Technology

- ASR enables virtual assistants to understand spoken requests

- Besides, background noise, length of utterance, and type of speech, virtual assistants' ASR performance degrades in presence of:
  - Nonnative or regional accented speech [6]
  - Stronger accents correlates with greater drops in accuracy [6]
- Native speech SOTA LID models degrades on non-native accents [6]

State of the Art Approaches

- SOTA LID models [6]
  - Resnet, XLS-R 300M
  - High error increase with nonnative accents
- Hybrid [6]
  - Acoustics + ASR-based
  - 50% relative error reduction
- Current SOTA [6]
  - Acoustics + character n-gram-based Naive Bayes (lexicon-free ASR Transcript representation)
  - Increase robustness to LID accented speech

Discussion

- Listening to mismatched native variants can lead to real-time processing costs in native listeners
- Non-invasive electroencephalography (EEG) shows promise for ASR using deep neural networks [11]
- Combining EEG representations with transformers could improve accent-robust LID [12]

References

[1] Lipski, 1994
[2] Bent & Bradlow, 2003
[3] Hayes-Harb et al.,, 2008
[4] Stibbard & Lee, 2006
[5] Goslin et al., 2013
[6] Kukk, K., & Alumäe, T. 2022
[7] Babu, Arun, et al. 2021
[8] Feng, Siyuan, et al. 2021
[9] Wu, Yunhan, et al. 2020
[10] Awasthi, Abhijeet, et al. 2021
[11] Krishna, G., 2021
[12] Hanani,et al., 2022