Investigating Speaker Diarization on Code-Switching Speech Executive Summary

Bronston Ashford z5146619

Justification for Thesis

Code-switching (CS) speech, the act of alternating between two or more languages in or between sentences, is a prevalent phenomenon in most multilingual societies [1, 2]. As the world becomes increasingly interconnected, speech technologies have an increasing demand to interface with these non-standard forms of speech. Among these technologies, speaker diarization (SD) systems, which identify 'who spoke when' in an audio clip, play a significant part.

SD systems serve as an important component in the pipeline of various speech technologies, such as speech recognition or meeting transcription. Since the quality of upstream SD directly impacts the performance of the parent technologies, audio recordings containing CS speech may be at a disadvantage compared to monolingual speech [2, 20]. Therefore, mitigating errors introduced by CS in audio would contribute to better accessibility and understanding in multilingual societies.

However, developing robust SD systems that can effectively handle CS speech requires an understanding of how much and to what extent CS speech impacts these systems. Yet, due to a notable lack of research in this domain, the question remains unexplored. Furthermore, contributing to the lack of research is the absence of freely available SD datasets containing CS speech. The quality of research on diarizing CS speech is intrinsically tied to the quality of available datasets, the generation of which can be challenging and time-consuming. Therefore, the addition of a free CS dataset to the field stands to facilitate future research and contribute to the development of this domain.

This thesis intends to contribute by initiating an investigation into the impact of CS speech on SD systems with a specific language pair. It aims to offer initial insights into how much SD systems are influenced by CS speech and to identify which aspects of CS speech present the greatest challenges. In doing so, a new CS dataset fit for diarization will be developed and made freely accessible.

Objectives

Literature Review

Preliminary Work

Plans

Meeting Log

Bibliography

[1] S. Sitaram, K. R. Chandu, S. K. Rallabandi, and A. W. Black. A Survey of Code-switched Speech and Language Processing. [Online]. Available: http://arxiv.org/abs/1904.00784

[2] T. J. Park, N. Kanda, D. Dimitriadis, K. J. Han, S. Watanabe, and S. Narayanan. A Review of Speaker Diarization: Recent Advances with Deep Learning. [Online]. Available: http://arxiv.org/abs/2101.09624