Kamila Barrios

James Hatch

Ana Leon Urrutia

Bias in Emotion Recognition: How Accent Shapes AI's Perception of Speech

With the rise of AI-powered resources for students entering the workforce, it makes sense to question the accuracy of the models that help students interview, upgrade their resumes, and even automate the application process. With interviews specifically, some AI assistants—including InterviewAI, AIApply, and Final Round AI—go as far as offering feedback on tone of voice, sentiment, and emotion, but just how accurate are they? Services like accent reduction software (such as Sanas and Krisp) make us question the extent to which voices are classified and perceived by others, and voice recognition software. Can we trust AI to correctly gauge the sentiment of different ethnicities and accents?

One pattern that we have seen among models is the adoption of widespread racial biases. While these issues are particularly prevalent in face recognition software (Scheuerman et al., 2021), we believe that implicit human biases have widespread effects on a variety of datasets and data models. One example of a sphere that might be influenced by such biases in human speech. While speech isn't directly related to race, global speech accents inhabit a similar sphere by indirectly conveying information about a person's linguistic origins. Ultimately, we wonder if accents experience the same biases that LLMs display to certain racial demographics.

To an extent, we believe that such patterns are likely to exist. According to a paper written by Jiang et al. (2019), the presence of an unfamiliar (outgroup) accent can alter the way in which humans interpret speech. In this paper, the believability of speech was affected by the presence of outgroup accents, with believability changing in relation to people's perception of the outgroup associated with the accent (Jiang et al., 2019). While we might not expect statistical models to draw conclusions about outgroups and compare this data with biases associated with the groups, a similar and much more realistic bias could arise from models not having enough minority group accent data. This pattern would tie into the unfamiliarity aspect of the Jiang et al. (2019) study.

These observations have driven us to question how accents alter results in certain voice recognition models. In particular, we identified emotion-recognition models as a point of interest due to the prevalence of training data and its direct relation to voice and vocal sounds. This caused us to raise the following questions: first, how do different emotion-classifying models respond to unfamiliar accents? Second, is there a bias for certain emotions in some ethnicities (over others) in emotion-classifying models? Finally, is there a bias for certain emotions in some ages (over others) in emotion-classifying models?

There are lots of possible datasets we can use to investigate these questions. One such dataset is the Toronto emotional speech set (https://www.kaggle.com/datasets/ejlok1/toronto-emotional-speech-set-tess), a set of 2800 audio files created by two actresses (ages 26 and 64). Recordings consist of audio clips of 200 target words, with each word receiving seven variations corresponding to different emotions (anger, disgust, fear, happiness, pleasant surprise, sadness, and neutral).

Another prominent dataset comes from Crema-D (https://www.kaggle.com/datasets/ejlok1/cremad/data). Crema-D is a dataset comprised of 7442 sound bites from 48 male and 43 female actors (ages 20-74). In each sound bite, one of 12 sentences is spoken using six different emotions (anger, disgust, fear, happiness, neutral, and sad) and four emotional levels (low, medium, high, and unspecified). Additionally, this dataset is collected from a variety of different races, including African American, Asian, Caucasian, and Hispanic.

A final dataset we've identified to be of interest is the Speech Accent Archive (https://www.kaggle.com/datasets/rtatman/speech-accent-archive). This dataset contains 2140 audio samples collected from English speakers. These speakers come from 177 countries and incorporate speakers of 214 different native languages. This dataset also contains a CSV dataset of all demographic data for each speaker.

We plan to use each spoken emotion dataset (like the CREMA-D and Toronto emotional speech datasets) to create emotion-recognition models for testing. While we hope to create some of our own emotion-identification models using the training data, we also plan to pull from some of the more popular models created for each dataset. We also predict that there will be instances where we combine data from multiple datasets at once. After doing this, we plan to introduce the resulting models to accent-based voice datasets (like the speech accent archive) and see how different accents are categorized emotionally. Using this information, we aim to create a new dataset that tracks the emotional scores and accent types of different audio files.
We then plan to examine these results and investigate for high concentrations of emotions and emotional groups (positive/negative) among different accents.

We expect to see models trained on emotion data attributing more negative emotions (anger, especially, but also sadness, disgust, etc.) to non-American accents (which we believe from the data is the predominant accent). We hypothesize that non-American accents will be categorized by emotion-categorizing models as being more negative due to either a lack of diverse data, resulting in unfamiliar-accent biases.

References

Jiang, X., Gossack-Keenan, K., & Pell, M. D. (2019). To believe or not to believe? How voice and accent information in speech alter listener impressions of trust. *Quarterly Journal of Experimental Psychology*, *73*(1). https://journals.sagepub.com/doi/full/10.1177/1747021819865833

Scheuerman, M. K., Hanna, A., & Denton, R. (2021). Do Datasets Have Politics? Disciplinary Values in Computer Vision Dataset Development. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 1-37. https://dl.acm.org/doi/10.1145/3476058

Work Agreement

Our aim as a team is to work together to find a workload division that makes sense for each of our interests and routines. We can acknowledge that an equal split of the workload might not be apparent at every stage of the project, but we want to work harmoniously in a way that will have split the work evenly by the end.