Alex Najera

Karen Mazidi

CS4395.001

13 November 2022

Beyond Laurel/Yanny: An Autoencoder-Enabled Search for Polyperceivable Audio

There were three authors involved in this paper: Kartik Chandra, Chuma Kabaghe, and Gregory Valiant. All from Stanford University, with Gregory Valiant being an assistant professor there.

As progress is made creating models that can get verbal speech as input and understand the words spoken, certain situations regarding a spoken language must be considered such as polyperceivability. The idea is that there are words or speech that can sound like one thing for one person, but something entirely different to someone else. A well known example is the “laurel/yanny” phenomenon, two different words were interpreted from one audio clip causing endless discussions online of what the clip was saying. And it brings to question how many polyperceivability examples exist in the English language.

Prior work has looked into specific examples of illusions that has helped systems catch such things, but as the paper stated, there has not been much prior work looking into “the extent to which ‘typical’ instances are polyperceivable”. Additionally, there has been work done such as Miller (1981) studying similar sounding words, but little in radically different words that can still be perceived in different ways in the same audio, such as yanny and laurel.

This paper was able to provide different examples found in a sample space that are polyperceivable, and using the sample space used to find such examples, also determined that at least 2% of English words, are capable to be polyperceivable. These illusions that can fool even people are important to be accounted for when determining the accuracy of machine learning systems.

The authors found this by acquiring a large amount of audio clips that were likely to be polyperceivable. Additionally, when finding polyperceivable examples in the sample space they used, they tested it with a crowdsourcing marketplace called Mechanical Turk to collect human responses and how they interpreted the clips. The selected clips were tested with the Mechanical Turk workers and found that there were a lot of split perceived sounds, finding that the clips were successfully found to be polyperceivable. The authors determined that the 2% estimation was conservative since the final samples they selected were selected manually by the authors.

Kartik Chandra has been cited 81 times in google scholar. Gregory Valiant has been cited by 4490, and Chuma Kabaghe had few results and didn’t show to have a specific number of citations. Gregory Valiant had by far the most citations as a professor. I believe this work is important because the ambiguity of how words sound in a spoken language needs to be understood to find the extreme outliers that a ML system would mislabel. Additionally, understanding how common polyperceivability is in a spoken language and how it can trick even native speakers could allow classifiers to be more refined and catch similar illusions that even humans could not catch. These capabilities would be of great use to parse and read audio clips to a greater extent.