Memo Audio Features

Karnkumar Dalmia

July 2020

1 Introduction

For our speech to knowledge graph algorithm, a key amendment was to not only learn a graphical representation based on words uttered, but also on audio features. For now, our team has plans to visually depict three audio features - stress, emotion, and volume..

2 Stress

For stress, we want to measure the amount of voice pressure applied at each word. In our knowledge graph, we plan to have a gradient value that varies linearly with the log pressure applied to voice of words. We can create a gradient around that gradient value to visually depict that component of the voice feature.

3 Emotion

Currently, there exist packages that predict one of "eight" distinct emotions. We can soft-max apply the return value of NN of each i^{th} emotion to get relative weight of each emotion. We also have a square color grid, where we can map the eight different emotions to eight different points on the color grid (4 midpoints, and 4 corners of the square). We then move linearly in each direction proportional to the relative weight in order to yield a color from the color grid. Hence, we can denote emotion of audio feature via color.

4 Volume

Lastly, for volume, we want to vary the size of the word bubble in the knowledge graph based on the volume. We can simple vary the diameter of the word bubble w.r.t the root mean squared value of the volume in the speech subsection (explained below).

5 Timestamp

In order for all three above to work, we need to, in our speech to knowledge graph, have start and end timestamps for each word uttered. Currently, the existing system only does speech to knowledge graph. We would need to amend a functionality to get start and end timestamps of each word.

6 Incorporating Timestamps

To incorporate timestamps, we need subsection our speech signals, and do all three analysis above with only that timed subsection of the speech for a given word..