Practical Exercise - Task 6
Clustering of Speech Signals
Speech Technology - COM4511/6511

May 2020

1 Introduction

A simple form of learning about inherent properties of speech data is unsupervised clustering. We will use K-means to cluster vowel and other speech sounds provided, and observe the outcome. The aim of this task is to increase understanding of the properties of speech sounds while learning about the specific properties an shortcomings of k-means clustering in the context of speech.

K-means clustering allows the grouping of vectors, on the basis of distance functions. Such grouping can then be analysed in the context of our knowledge on speech production mechanisms. For interpretability of machine learning outcomes it is desirable if automatic data organisation methods yield similar results to what we know about speech signals.

2 Preparation

From the MOLE module website download vowel (and other) sounds and helper scripts. An MFCC computation script is provided to you as a package called "speechtech".

For this task we will use "scikit-learn", a Machine Learning toolkit for Python. First, you need to install "scikit-learn" if you haven't done so. It is recommended to also install a Python visualisation library called "seaborn".

http://scikit-learn.org/stable/install.html

```
sudo pip install -U scikit-learn
sudo pip install seaborn
```

In order to get started, take the following first steps:

1. Unzip "vowels.zip" and place the "vowels" folder that contains 4 vowel sounds in the same folder where you put the helper script "speechtech_kmeans.py" . Note that other sound file collections are provided.

2. Place the package "speechtech" in the same working folder.

3. Provided the package "scikit-learn" is installed, you should be able to run the "speechtech_kmeans.py" helper script and see some results. Word through the code.

3 Objectives

The objective in this task is to

- Explore the outcomes of k-means clustering of speech data.

- Implement your own k-means clustering to explore the use of different distance functions, initialisation methods and feature representations.

- Explore the outcomes as a function of use of different data

4   Desired outcome

The outcome of this task is expected to be

1. Source code

   The helper function extended with your own code, possibly multiple versions of the function with different implementations. The code should be reasonably documented and a clear reference to relevant code should be included in any discussions in the report.

2. A brief report

   Report your findings with k-means clustering of vowels and other speech data. You should provide an analysis of the results, and how these results could be altered to improve their meaningfulness.

   In particular you should explore different distance functions, include those that include concepts of temporal modelling.

   You can also record your own voice and take fragments of your own speech (vowels, consonants or other groupings) for analysis. If you record your own speech, an easy tool to use is audacity (https://www.audacityteam.org/).