**Definition of formants**

In speech science, a formant is defined as the spectral shaping that results from an acoustic resonance of the human vocal tract.

Formants are often measured as amplitude peaks in the frequency spectrum of the sound by a spectrogram and, in the case of the voice, this gives an estimate of the vocal tract resonances.

**Our dataset**

Our dataset was taken from the emuR package of R.

In particular, it is called dip and it contains a total of 186 samples of three diphthongs (aI, aU and OY) recorded by two native German speakers (one male and one female).

A diphtong is a fusion of two vowels (e.g. right -> ‘aI’).

Each sample has a specific duration with an overall average of 29.4 time units (1 time unit = 5 ms) and it contains several information, the most important being the frequency values of the first 4 formants.

Other relevant information are the labels for the diphtongs and the speakers.

Each sample is composed by several observations, one for each time unit, for a total of 5462 observations.

**State of the art**

Many studies have been carried out about the formants frequency values for both vowels and consonants.

We used these studies, in particular the ones involving the vowels, in order to focus on the most important features of the dataset.

One of the most important results of these studies is that the first two formants are generally sufficient in order to characterize the vowels.

Another important result that can be useful to recognize vowels is the ratio F2/F1 (and F2-F1).

**Problems encountered**

The original dataset included null formant frequences, obviously due to wrong measurements, therefore we took care of them by replacing them with the mean value of the correspondent pair diphtong, speaker of the same formant.

(Although the original dataset presented some interesting methods it was not so easy to manipulate as the access to the data (contained in dip$data) was kind of tedious. Therefore, we manually created some dataframes containing relevant information.) probabilmente sarebbe meglio non dirlo

**What we have done so far**

*Overall analysis of the dataset*

Due to the different length of the samples we needed to “normalize” the frequency values over time, in order to be able to compare them. Therefore, we computed the values of all the samples for each formant and we kept 11 of them (from 0% to 100% of the sample length).

Our main focus has been to analyze the differences among diphtongs for the same speaker since, as we said before, the differences between males and females are very pronounced and therefore less interesting.

However, we are still performing analysis on the differences between speakers since this may lead us to interesting results for the diphtong classification problem.

*PCA*

*Gaussianity*