Speaker classification project

Signal processing

Dell'Aria Doriano **Goffaux Lionel**

14 novembre 2020
Academic Year 2020-2021
Bachelor's degree in Computer Science

1 Introduction

In this project, we are going to extract features of a set of voice samples, to be able to classify if the sample is a male or female voice.

2 Energy

In the Figure 1, we plot the sample and its energy. We observe that with a threshold of 5, we are able to classify the voiced and unvoiced frames. Then we use these voiced frames to extract all the features we need.

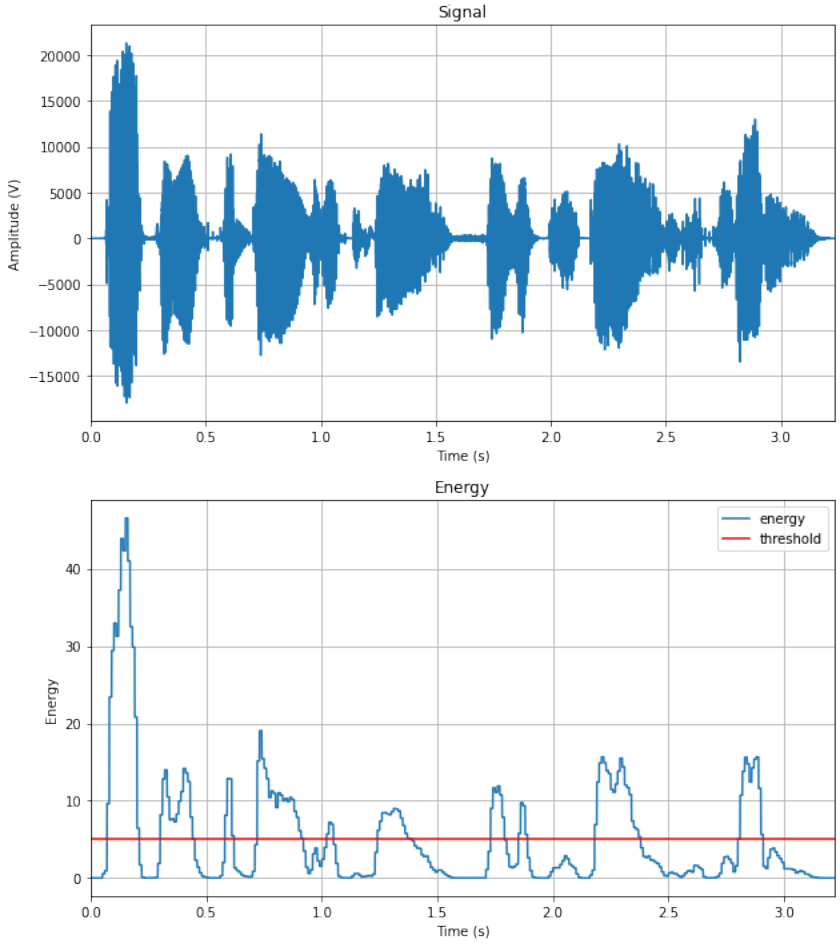

FIGURE 1 – Energy of a sample

3 Features extraction

3.1 Pitch

They are multiple ways to exctract the pitch of a sample. In our case, we extract the pitch with an autocorrelation-based system and a cepstrum-based system.

3.1.1 autocorrelation

The principle of autocorrelation is to correlate the signal with itself. We first split the signal into frames then autocorrelate frames with themself. So we obtain the pitch by measuring the distance between the distance of two peaks of the autocorrelation. we fix the frames width = 21, the step = 5 and the threshold = 5. In the Figure 2, we observe that the method doesn't have 100% accuracy because we obtain pitch over 5000 Hz.

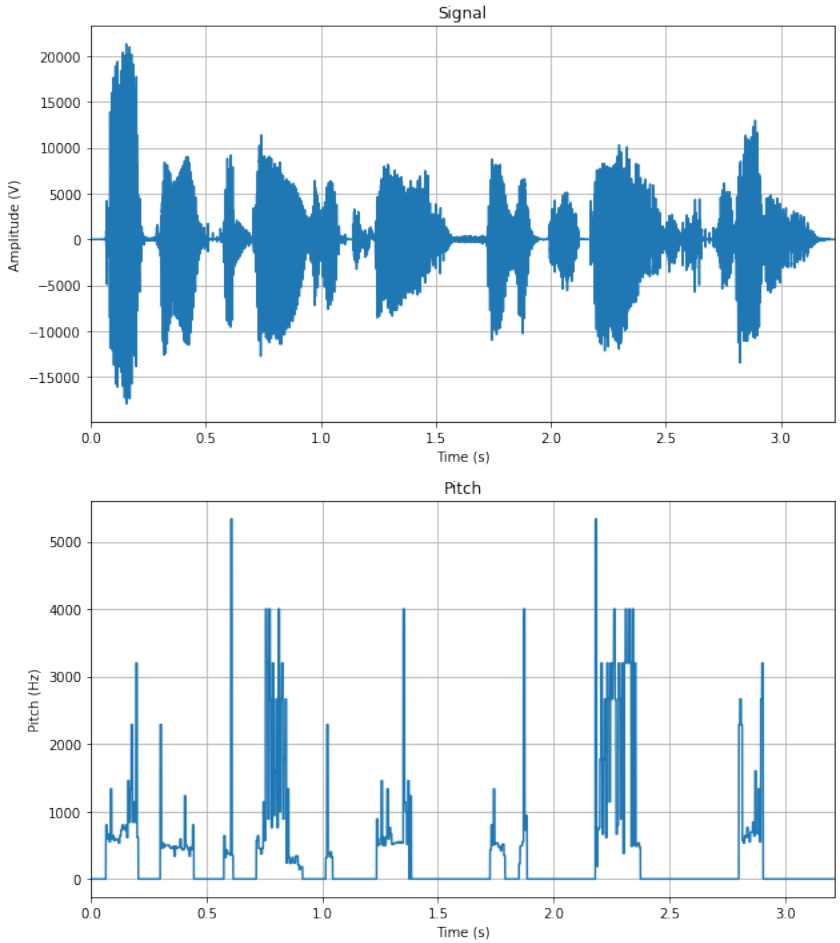

FIGURE 2 – autocorrelation-based system

3.1.2 Cepstrum

test

3.2 Formants

3.3 MFCC

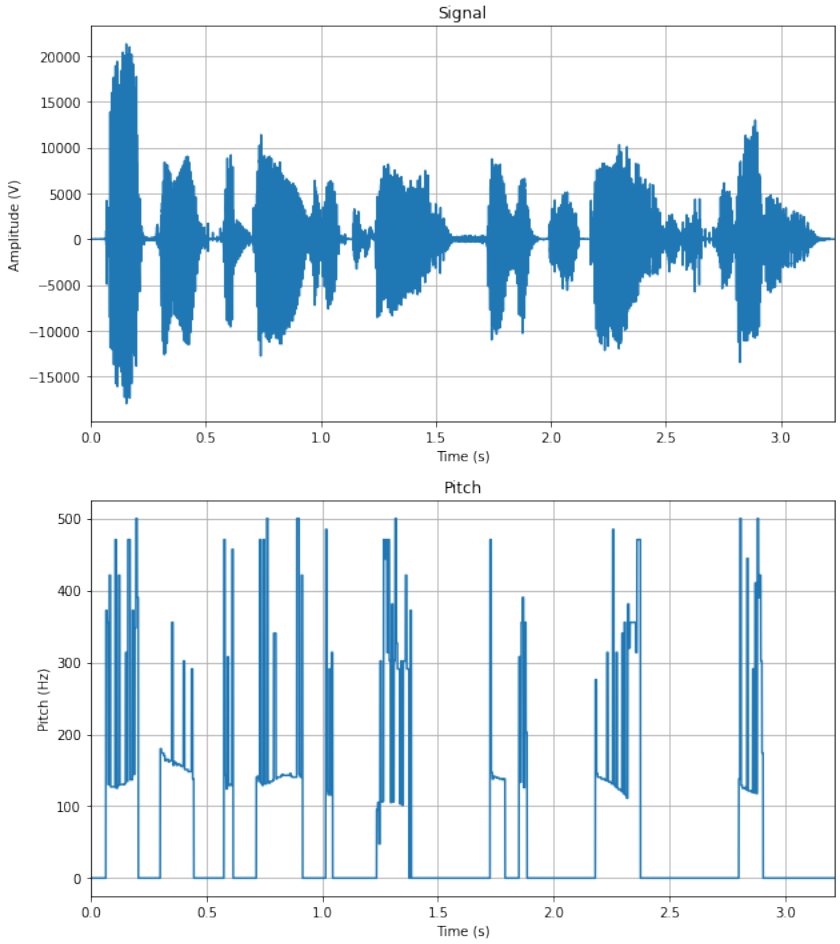

FIGURE 3 – cepstrum-based system