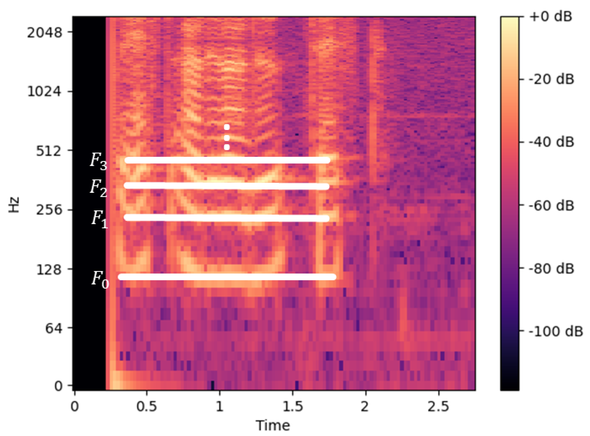
Modifying the Voice of a Recording

Antonin CHIRON, Valentin HAMELIN, Jules TOUPIN

INSA CENTRE VAL DE LOIRE

Frequency Characteristics of Human Voice

Before attempting to modify the voice, let's first look at its processing.
The human voice produces a fundamental frequency (F_0) between 80Hz and 400Hz, which depends on the age and gender of the person. Like any instrument, the voice also produces several harmonics (F_1, F_2, ...).
The timbre of the voice is characterized by the frequency of the fundamental and its harmonics. We can observe the fundamental and harmonics on a time-frequency representation, depicted below.



Here, and for the rest of the study, the intensity scale is arbitrary and does not represent a meaningful measurement or magnitude of the sound intensities involved in the recording.

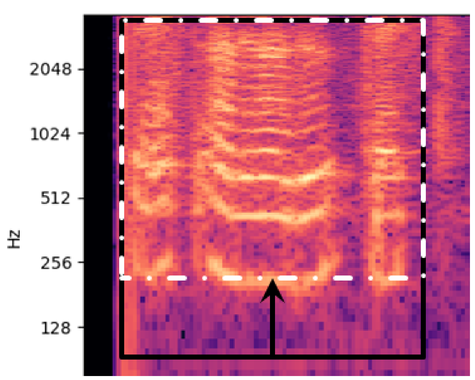
The figure above shows the time-frequency diagram of a recording with a voice and no other significant sound recorded.

Modifying Voice Timbre

o perform the following modifications, we will use the Librosa library in Python, which will facilitate the manipulation of the recording in the frequency domain. We will specifically use the command: `librosa.effects.pitch_shift`.

Making the Voice Higher:

The higher the frequency of a sound, the higher it is perceived. One could isolate the fundamental with an algorithm detecting F_0 and only shift it to higher frequencies. However, this results in a very synthetic-sounding voice. To achieve a satisfactory result, we will instead multiply all frequencies by a more or less significant factor. Graphically, this looks like a compression of the spectrum towards higher frequencies.

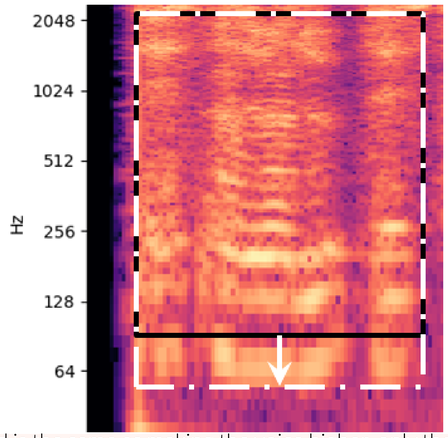


This method has the advantage of shifting all harmonics with the fundamental, making the result sound more natural and less synthetic, as long as the modification is reasonable.

ote: With this method, all sounds in the recording are modified, not just the voice. This means that every audible sound in the recording will be higher-pitched.

Making the Voice Deeper:

Similar to making the voice higher, we will multiply each frequency by a variable factor to make them lower. Graphically, this may appear as an extension of the spectrum towards lower frequencies.

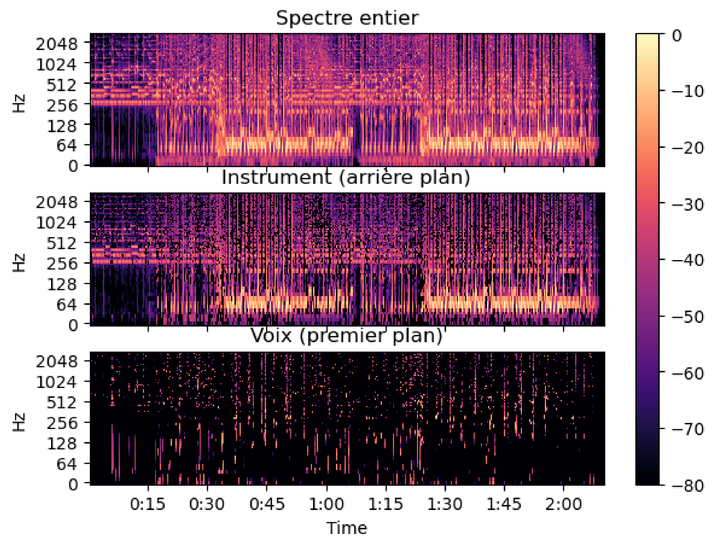


ote: This method is the same as making the voice higher, only the result changes. It has the same drawbacks.

Example of Use on Music with Instruments

Separation of the Voice from the Recording:

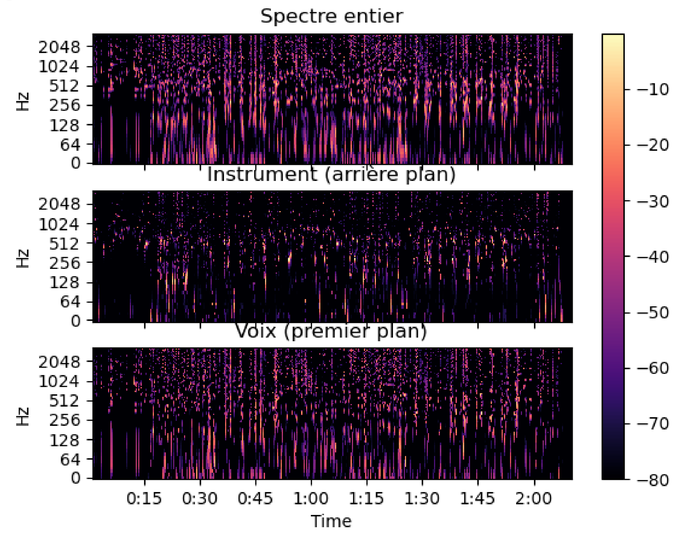
The methods seen previously apply to the entire recording, not just the voice. Therefore, we must proceed with separating the voice from the original audio. For this, we will use a method that employs "soft mask" filters (which we will not detail here), which has the advantage of taking into account the probabilistic nature of the signal, resulting in a more satisfactory outcome than a simple band-pass filter, which makes a much coarser cut.



ote: It is much harder to identify the fundamental and its harmonics in the recording we are processing here because there are multiple voices and the recording is much larger than the previous example.

Additional Voice Separation: (optional)

If the initial voice separation is not satisfactory (audible instruments in the background), we can repeat the separation process to better isolate the voice. This gives us:



Final Result:

Once the voice has been properly separated, one of the two methods mentioned earlier must be applied depending on the desired outcome. This result is then added to the previously extracted background.
The QR code below leads to a Box folder (an alternative to Google Drive) where there are two examples of what can be achieved by modifying the voice as described.

SCAN ME