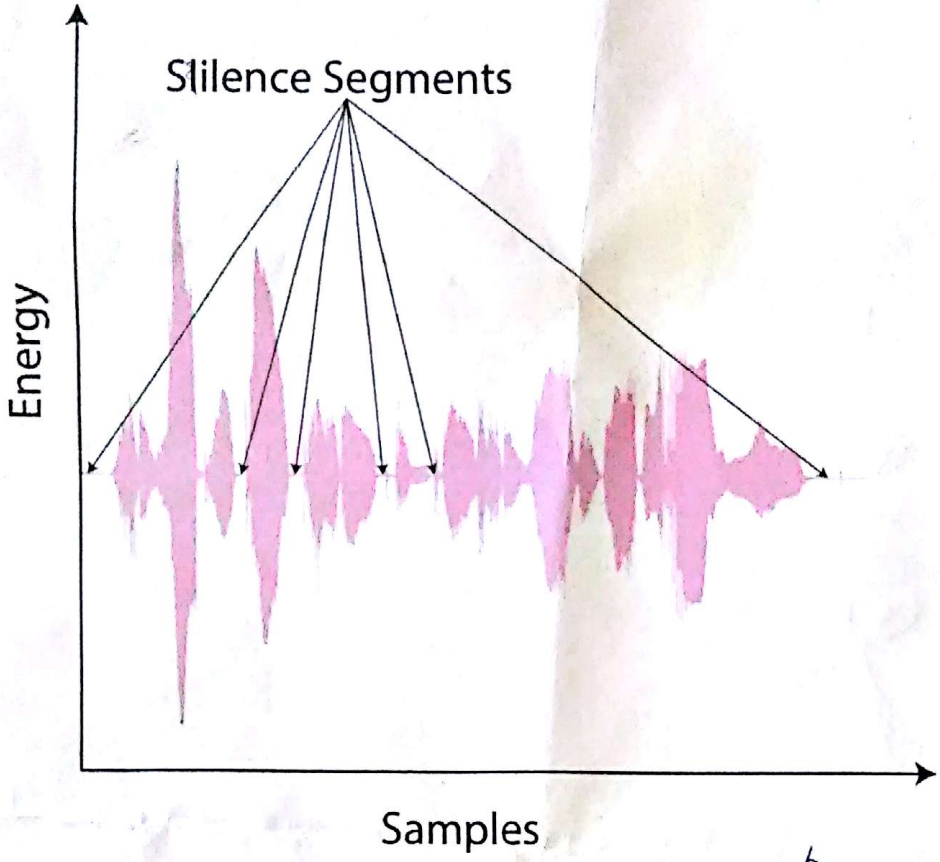
A speech signal is comprised of energy sampled at different time points. Each signal holds voiced and unvoiced segments. The goal here is to extract the voiced segments.



184000 samples

You are given a path to a file containing speech.

nst.wav 23 s,
8 kHz sampling rate ~~44.1 kHz~~
~~1323000~~ samples

1. Load the file and extract the speech signal as a numeric vector of samples and the sampling rate
2. Calculate the mean (μ) and standard deviation (σ) of the first 200ms samples.
3. Go from the first sample to the last sample. For each sample check whether $\frac{|x-\mu|}{\sigma} <$ 3. If this is true, mark the sample as 0. Otherwise, mark it as 1.
4. Divide the entire signal into 10ms non-overlapping windows.
5. In each window, there are M zeros and N ones. If M > N, convert all ones to zeros. Otherwise, convert all zeros to ones.
6. Collect samples from the original signal corresponding to 1 from the converted signal.

Costomer

dudu@multisense.nl