ACOUSTIC EFFECTS OF VARIOUS FACE MASKS ON SPEECH SIGNALS

CT 437 Speech Communication

Name: Jitanshu Shaw
ID: 201901292

20th December 2022
DAIICT, Gandhinagar

Abstract

Face masks stifle speech and make it more challenging to communicate, especially for those who have hearing loss. This study uses a built-in microphone for recording and getting the .wav file and a live human talker to the speech signal to compute and analyse the difference between masked and No mask speech signals. The mask that I used in this project is two layered cloth.

Introduction

- How a different types of masks effects on speech signals?
- What type of clothes or materials is used?
- What are the differences between masked and no-mask speech with four feature extraction methods?

Method

- Collecting data
- Feature extraction
- Plotting graph
- Comparing both the graph

Dataset

12 female and 5 males(self-collected)
Sentence: “The quick brown fox jumps over the lazy dog”

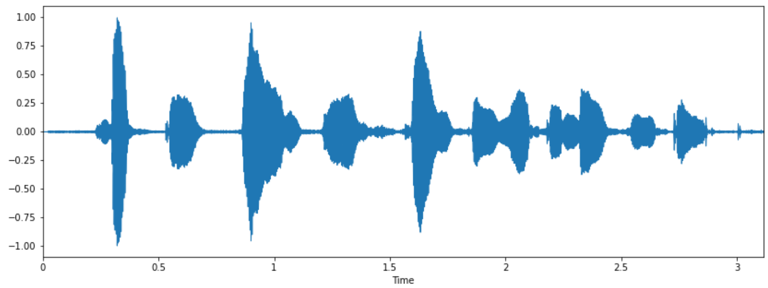


Fig 1: Speech Signal without a mask

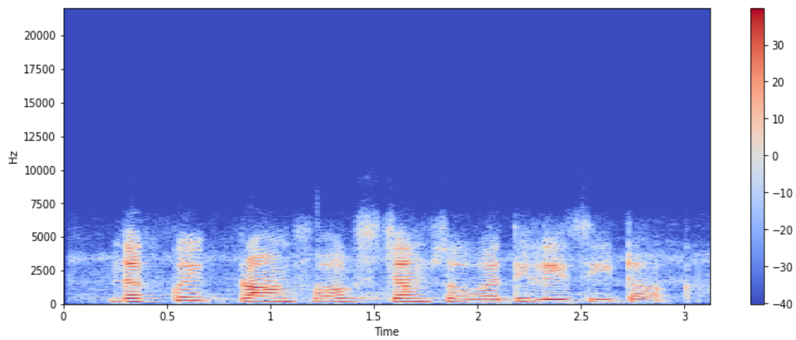


Fig 1.1: Narrowband spectrogram of the fig 1 speech signal

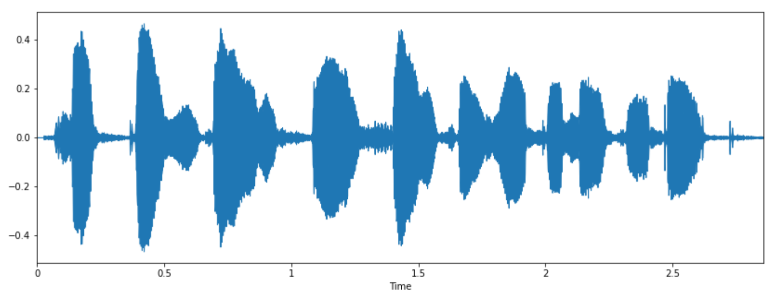


Fig 2: Speech signal with cloth mask

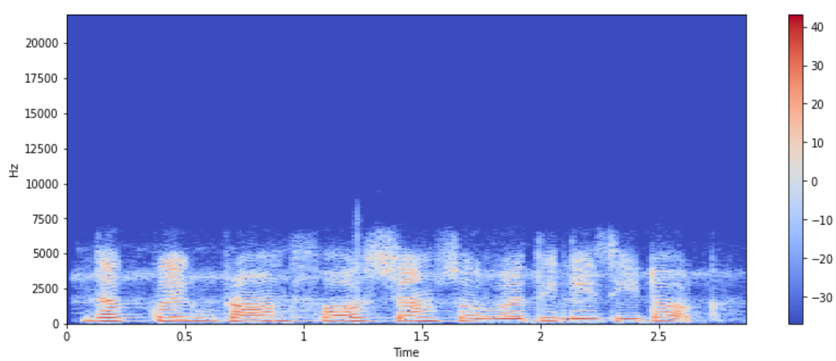
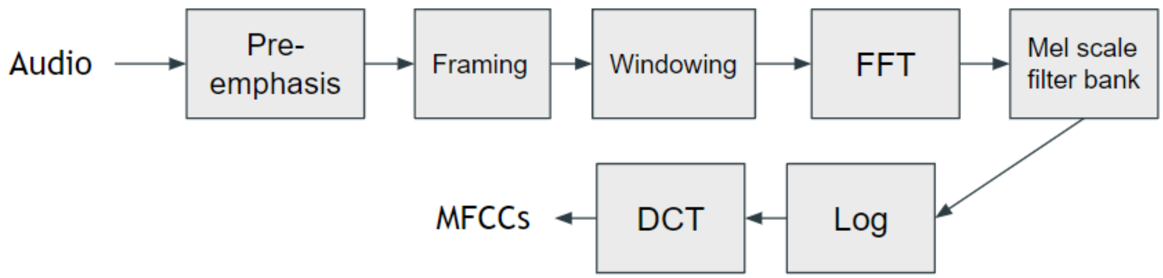


Fig 2.1: Narrowband spectrogram of the fig 2 speech signal

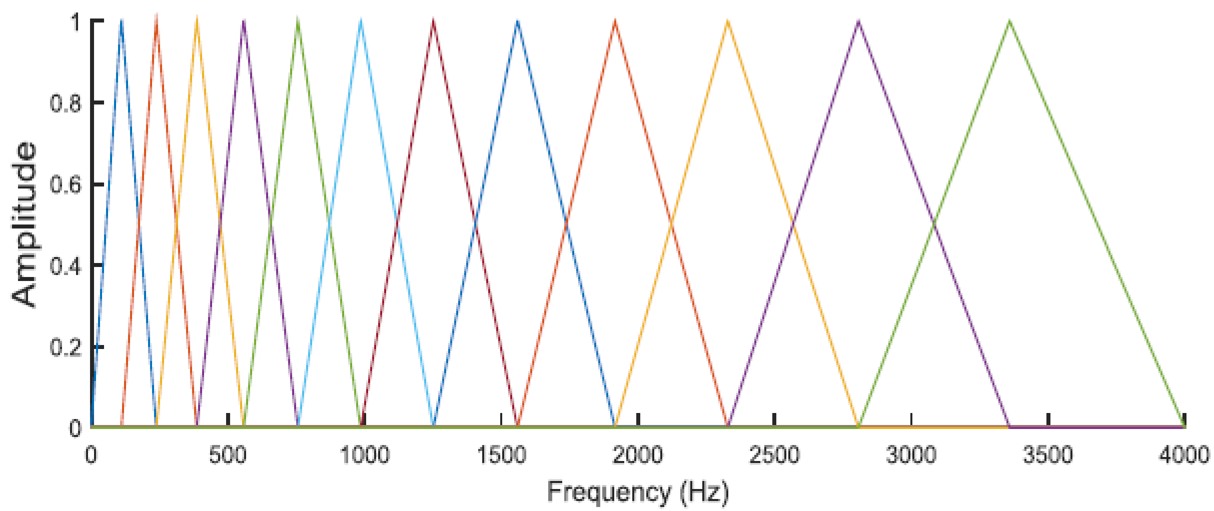
From comparison between fig 1 and 2 the signal in fig 1 is much sharper in compare to fig 2 signals. The gap between the any two words in fig 1 is much clearer than in fig 2 .

From comparison between fig 1.1 and 2.1 we can see that the redness in the fig 1.1 is much higher than in fig 2.1 which shows the energy of fig 1.1 is higher than in the spectrogram of ‘cloth mask’ speech signal as the voice is muffled so we can see that the energy is lost in fig 2.1

Block diagram for MFCCs(Mel-frequency cepstral coefficients)



Mel scale filter bank



In order to mimic a Mel scale, a set of triangular filters are utilised in this picture to compute a weighted sum of the filter's spectral components. The magnitude frequency response of each filter is triangular in shape, equal at its centre frequency, and decreases linearly to zero at the centres of two consecutive filters. The result of every filter is then the total of all filtered spectral components. The Mel is then calculated using the following equation for a given frequency f in HZ: $F(Mel) = [2595 * log\,10\{1 + (f/700)\}]$

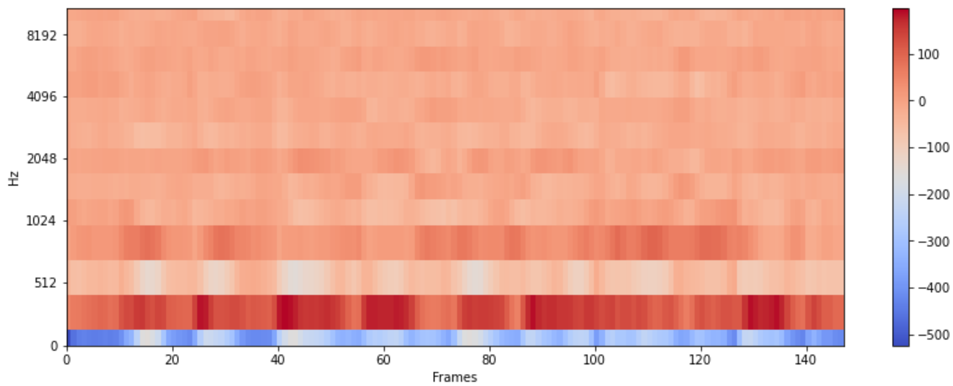


Fig 3: Shows the no mask MFCCs

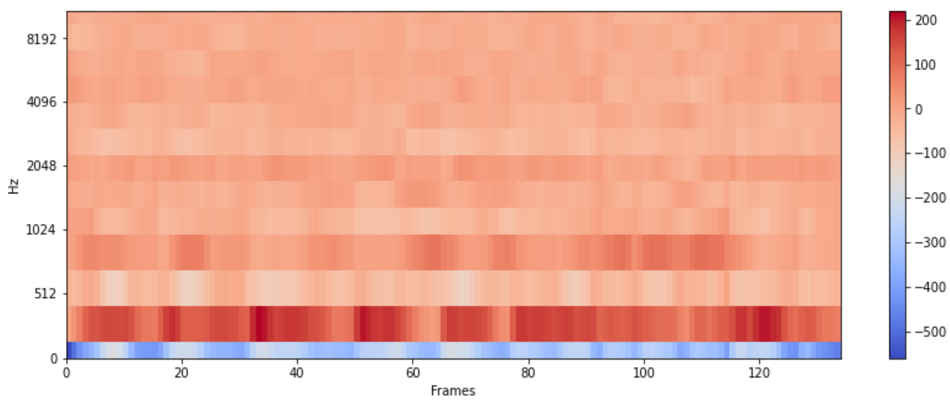
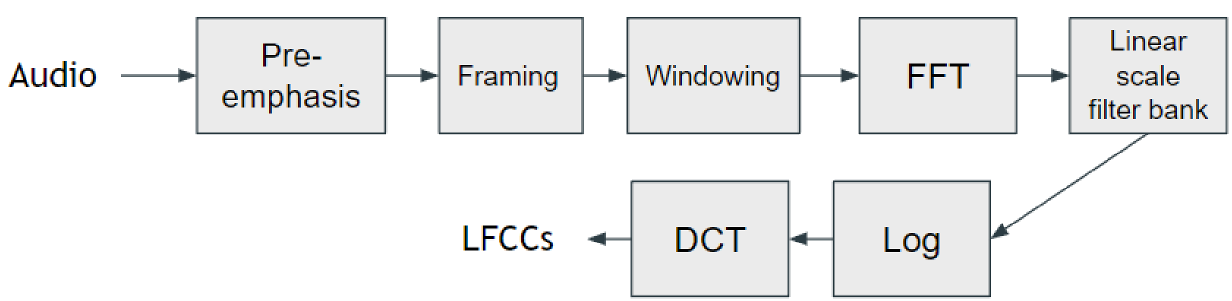


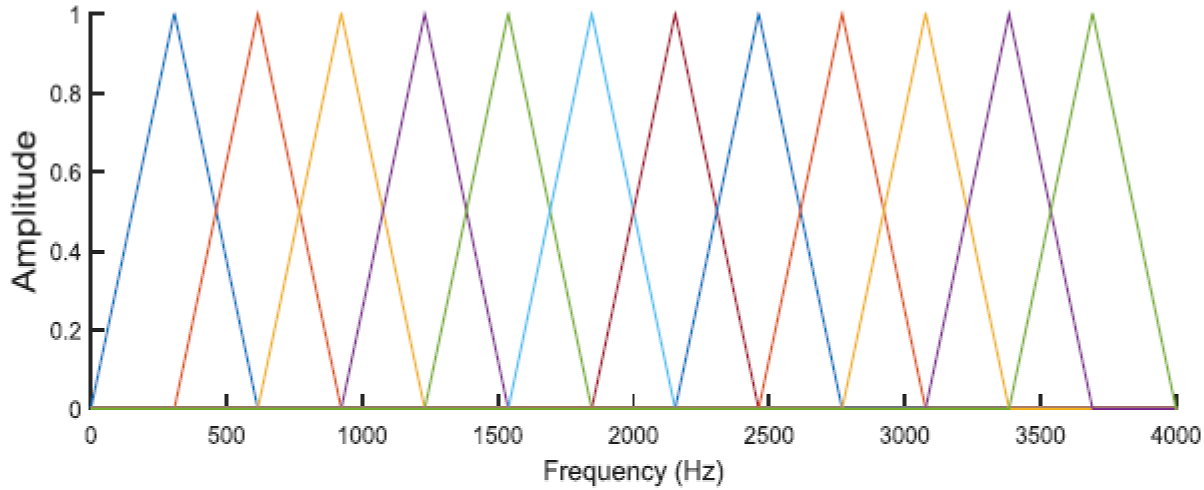
Fig 3.1: Shows the cloth mask MFCCs

Above (fig 3-3.1) is divided into 13 coefficients. The energy in fig 3.1 in comparison to fig 3 has lost much more energy in each band than fig 3, and the white colour is much more discrete in fig 3 than in fig 3.1.

Block diagram for LFCCs(linear frequency cepstral coefficients)



Linear scale filter bank



steps for calculating LFCC The signal is split into frames in the first stage, typically using a rectangular windowing function with overlap at predetermined intervals. As a result, each frame may be thought of as a vector of cepstral features. The discrete Fourier Transform is then applied to each frame, followed by the application of a triangle filter with evenly spaced frequency bins. On each output energy of each triangular filter, the logarithm is calculated. The Discrete Cosine Transform is then used to decorrelate the components. The benefit of this is that it lowers the overall amount of features in each vector.

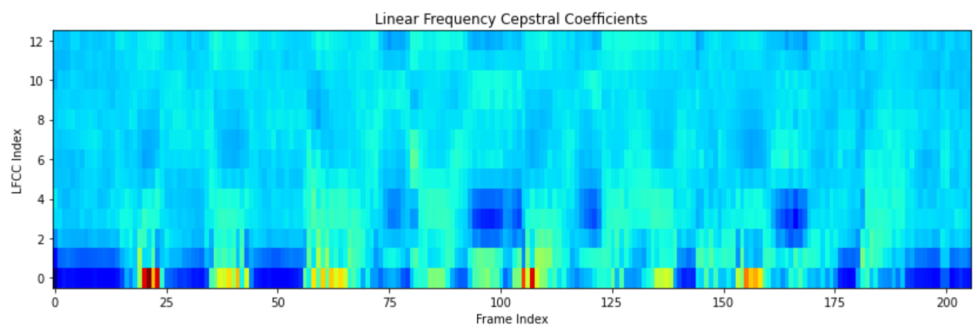


Fig 4: Shows the no mask LFCCs

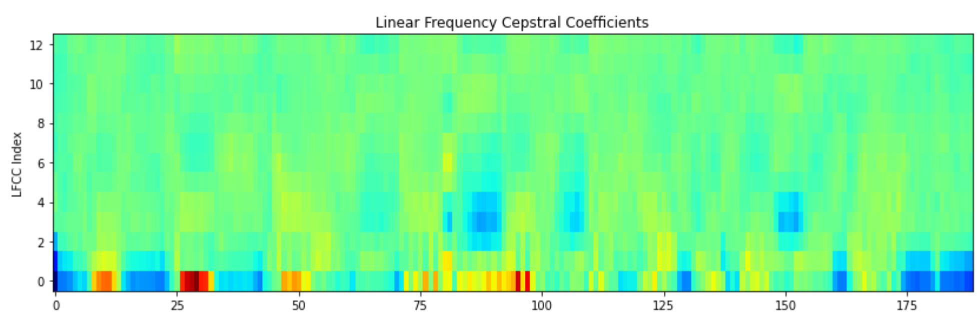
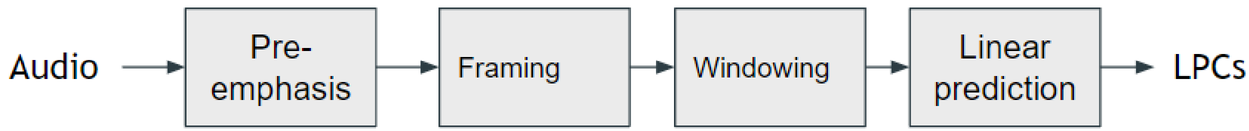


Fig 4.1: Shows the cloth mask LFCCs

The energy in fig 4.1 in comparison to fig 4 has lost much more energy in each band than fig 4, and the blue colour is much denser in fig 4 than in fig 4.1.

Block diagram for LPCs(Linear Prediction Coefficient)



LPC(Linear Predictive coefficient):

The idea of linear prediction states that one can roughly approximate the current speech pattern as a linear aggregate of the previous samples. Because of its speed and precision, the LPC approach is essential in speech evaluation. It provides a great estimate of the vocal tract spectral envelope. LPC is used to calculate the characteristic vectors across each frame. Depending on the speech sample, application, and range of poles within the version

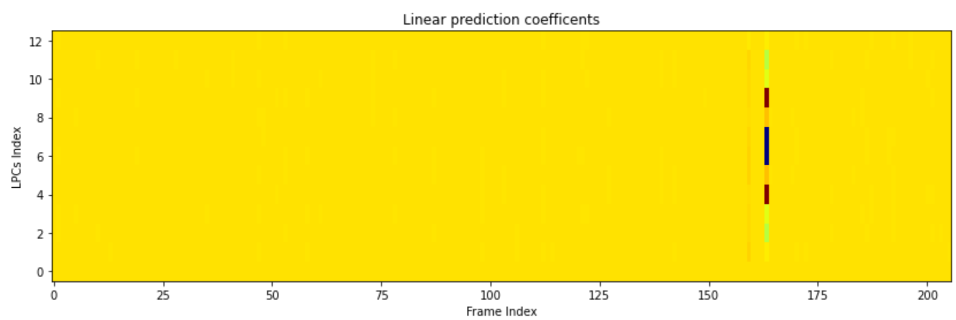


Fig 5: Shows the no mask LPCs

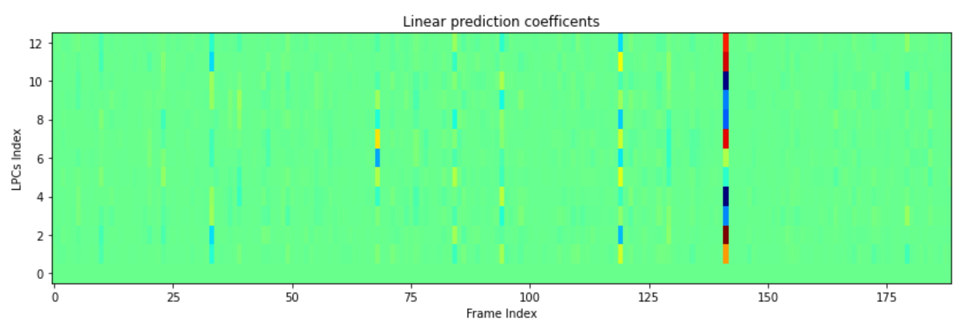
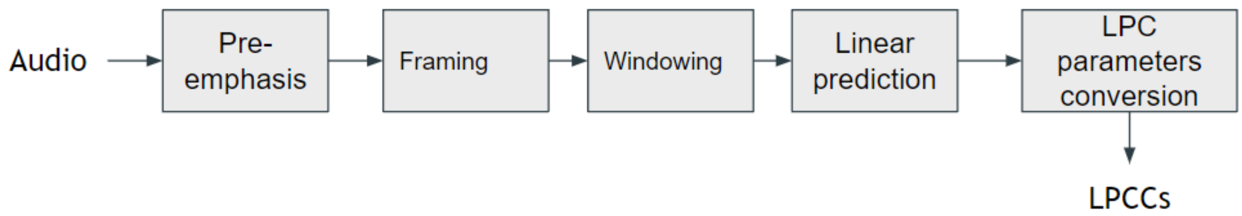


Fig 5.1: Shows the no mask LPCs

Block diagram for LPCCs(Linear Prediction Cepstral Coefficients)



LPCCs(Linear Prediction Cepstral Coefficients):

LPCC has emerged as one of the most popular algorithms for evaluating the fundamental properties of a speech signal. In order to improve the accuracy and robustness of the voice features extracted, the LPCC technique combines LP and cepstral analysis by taking the inverse Fourier transform of the log magnitude of the LPC spectrum. The basic idea behind this method is that one speech sample at the present time can be predicted as a linear combination of past speech samples.

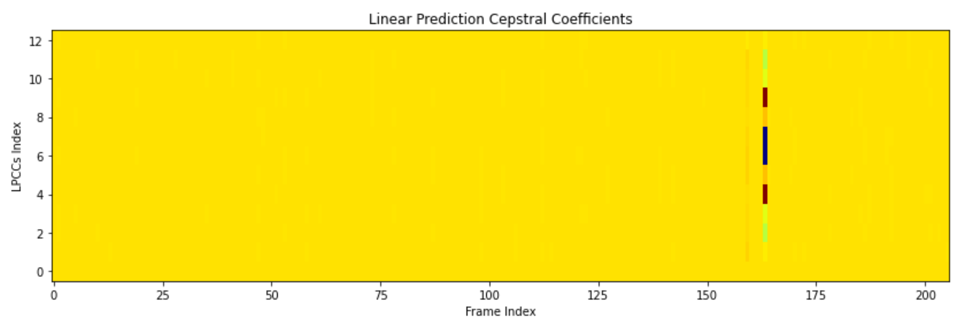


Fig 6: Shows the no mask LPCCs

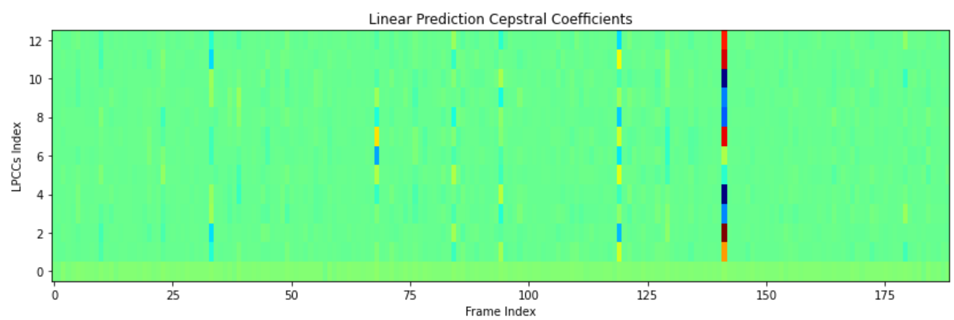


Fig 6.1: Shows the no mask LPCCs

Result

These experiments show that face masks attenuate high-frequency sound in front of the talker, and when a person with hearing loss listens to the word like “ma” “pa” and “ba” or any word that contains these syllable(“ta” “ka”) will have a hard time understanding what the speaker wants to tell.

References

- Ryan M. Corey, Uriah Jones, and Andrew C. Singer, "Acoustic effects of medical, cloth, and transparent face masks on speech signals", The Journal of the Acoustical Society of America 148, 2371-2375 (2020) https://doi.org/10.1121/10.0002279 last accessed:03-12-2022(13:13)
- Jung, Youngmoon & Kim, Younggwan & Lim, Hyungjun & Kim, Hoirin. (2017). Linear-scale filter bank for deep neural network-based voice activity detection. 1-5. 10.1109/ICSDA.2017.8384446. Last accessed: 20-12-2022(05:56)
- Yadav, M. S., & Kumar Jain, A. P. (2017, December). MFCC AND LPCC ANALYSIS OF SPEECH SIGNAL FOR DIFFERENT LANGUAGES. https://www.jetir.org Retrieved December 20, 2022, from https://www.jetir.org/papers/JETIR1712146.pdf
- Sehili, Amine & Istrate, Dan & Boudy, Jérôme. (2022). Primary Investigation of Sound Recognition for a domotic application using Support Vector. Last accessed: 20-12-2022(19:56)
- Jeremy I. Skipper, Virginie van Wassenhove, Howard C. Nusbaum, Steven L. Small, Hearing Lips and Seeing Voices: How Cortical Areas Supporting Speech Production Mediate Audiovisual Speech Perception, Cerebral Cortex, Volume 17, Issue 10, October 2007, Pages 2387–2399, https://doi.org/10.1093/cercor/bhl147 Last accessed: 20-12-2022(19:56)