

Pitch And Frequency Analysis of Transgender Voice

Keerthana A

Electronics and Communication Engineering Dept.
PES University, Bangalore, India

Veena S

Guide
Electronics and Communication Engineering Dept.
PES University, Bangalore, India

Abstract:- The paper deals with the conversion of masculine voice of transgender to feminine voice. This conversion of voice is carried out by manipulating the two major acoustic features of speech signal. The two acoustic features here are the frequency and pitch. Frequency is a physical quantity which gives number of vibrations per second whereas pitch is a perceptual quantity which depends upon the listener. In fact our ears detects the pitch not the frequency. The phase 1 of the paper, software execution was carried out in the MATLAB software. The first step would be obtaining the input signals. The input signal is a recorded audio signal in the .wav format. Next step is to analyse and detect the pitch of the input speech signal. The pitch analysis is done by the auto-correlation method. The low pass filter is used to get rid of the high frequency noise signal in the form of impulse. Then the input signal decomposed into number of windows depending upon the total length of the signal. The decomposition and overlapping of the signal is the PSOLA method. By using the PSOLA method we increase the frequency of the input signal to get the desired output signal. The overlapping of signal increases the frequency of the speech signal. Thus the desired output is generated by manipulating the recorded input signal. The hardware execution part has a lots to do with the real-time implementation of the paper. The hardware is required to obtain the input signal in real-time and give out the output in real time with a small delay. Here in this paper the input signal the raspberry pi 4 model B is used to obtain the input through a microphone and give out the output through the Bluetooth speaker. The raspberry pi 4 model B follows the stand-alone execution , to give out the desired output in real-time.

Keywords: MABLAB, voice modification, transgender, auto-correlation, PSOLA technique.

1. INTRODUCTION

The main aim of the paper is to come up with implementation that can analysis the pitch and frequency of transgender to convert the masculine voice of the transgender person to feminine voice. One of the major challenges faced by them is because of their voice. The people of this community are identified by their masculine voice. They are also being bullied and treated differently because of their voice in many situations. So the main aim of the paper is to find means to modify their voice in real time and make their voice a shriller one like other females. This paper deals with the two major acoustic features of speech signal. The two

acoustic features are pitch and frequency of the speech signal. Frequency is a physical quantity which gives number of vibrations per second whereas pitch is a perceptual quantity which depends upon the listener, in fact our ears detects the pitch not the frequency. In general, the pitch frequency of female is higher than that of male. This is because of the structure of the male and female vocal cord. The vocal cords are two bands of smooth muscle tissue found in the larynx (voice box). The length of male vocal cords is longer while the vocal cords of female are shorter. Since the length of female vocal cords are shorter the frequency of speech signal is higher than the frequency of speech signal that is delivered by male. In the implementation of analysis, the input transgender voice signal is manipulated based the acoustic features and the output signal is generated which is the feminine voice. The first phase dealt with the software simulation of the paper on the MATLAB software. In the software execution part, the input signal is obtained and the pitch is is detected by using auto-correlation technique. The impulse noise in the input signal is removed by using a low pass filter. Later by using PSOLA technique the signal is decomposed into windows and the windows overlapped to give out high frequency output signal. Finally the desired output signal is generated. The hardware execution part has a lots to do with the real-time implementation of the paper. The hardware is required to obtain the input signal in real-time and give out the output in real time with a small delay. The raspberry pi 4 model B is used to obtain the input through a microphone and give out the output through the Bluetooth speaker.

2. METHODOLOGY

For the implementation of voice transformation, the acoustic features of the voice signal have to be modified. This paper deals with the major acoustics features chosen, Pitch and Frequency. This modification procedure carried out in four major steps

- i. Obtaining input signal and filtering
- ii. Analysing and detecting
- iii. Decomposition to create windows
- iv. Overlapping and output generation.

2.1 Obtaining input signal

The first step would be obtaining the input signals. The voice of a transgender person is recorded and stored in the system. The format of the recorded audio file is noted. The audio is recorded in a silent environment and we consider to be a noise free signal or with negligible noise. Here for this application the signals frequencies above 900 Hz are eliminated by using low pass filter. Thus the recorded input signal is exempted from noises.

2.2 Analysing and detecting

The input signal is analysed and the acoustic features are detected. Since the pitch frequency is analysed. The pitch period is detected using the auto-correlation method. The signal is normalised. The normalisation process is adapted to smoothen the input signal.

2.3 Decomposition to create windows

The frequency pitch detected signal is then decomposed to create windows. The number of windows is based on the total length of the input signal. The windows created are hanning window.

2.4 Overlapping and output generation

The windows in the form is decomposed signal is overlapped. The overlapped is done by following the PSOLA technique. The signal overlapping increases the signal pitch frequency. The desired transformed signal is obtained by following all the above steps. Finally the generated output signal is saved in the system. The output signal generated is in the .wav format. The run command will perform all the steps and play the generated output.

Thus by following the steps above, the aim is achieved. That is low frequency input signal is transformed into high frequency output signal to achieve the goal. The figure 2.1 describes the flow of the signal processing

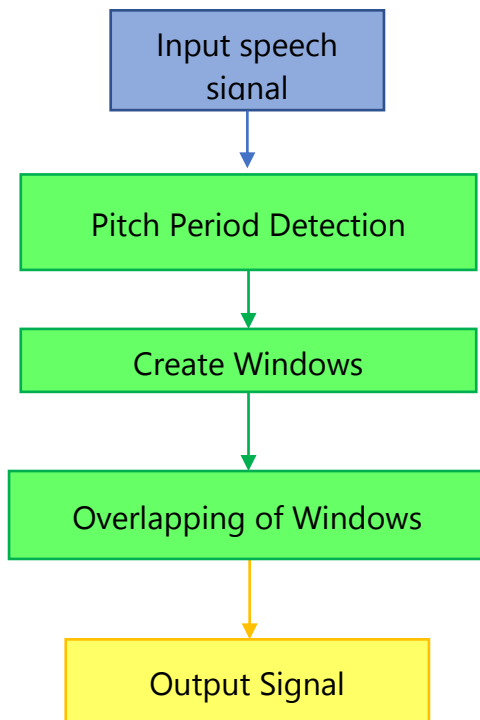


Fig: 2.1: Flowchart

3. IMPLEMENTATION DETAILS

3.1 Auto-correlation technique

The auto-correlation technique is used for the detection of the pitch period of the input signal. The frequency of the speech signal usually varies between 40 Hz for a low-pitched speech signal to 600 Hz for a high-pitched signal. This auto-correlation technique requires two pitch periods

to detect the pitch of the signal. That is nothing but, if we need to detect the low-pitched frequency signal, we will need to analyse at least 50 ms (milliseconds) of the speech signal. So, in this 50 ms time period the speech signal is high frequency need not necessarily have the same high frequency throughout the window.

3.2 Pitch synchronous overlap add (PSOLA) technique

Pitch synchronous overlap add technique is used to increase the frequency of the input signal. The PSOLA technique is usually used in the speech processing application, especially in the synthesis of speech signal in digital signal processing. This technique is usually adapted for modifying the duration and pitch of the speech signal. In the PSOLA technique the speech signal is decomposed into windows. The pitch modification and the duration modification is achieved by different styles of overlapping of signals. That is to increase the pitch of the speech signal, the windows are brought together, whereas to decrease the pitch, the windows are moved far apart. Similarly, to change the duration the segments are repeated several times to increase duration and few segments are eliminated to decrease the duration. Here for this paper we bring the windows together to increase the frequency. Later the segments are combined together by the overlap add technique. This is how PSOLA technique works to increase the frequency.

4. RESULT

The desired output is a high frequency output signal. The figure 5.1 is the figure obtained as output after the execution of the MATLAB code.

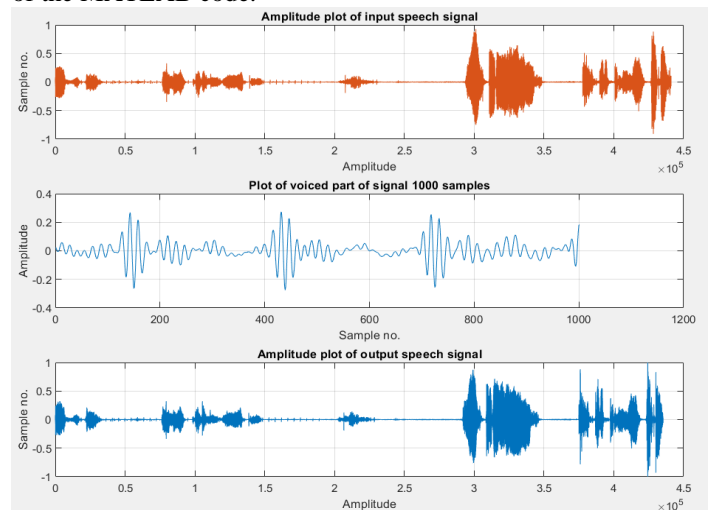


Fig: 5.1: Simulation results

We have three different graphs in the figure. The first one on the top is the amplitude plot of the input speech signal. For demonstration of the autocorrelation technique, we have used 1000 samples from the input signal. The last graph in the figure at the bottom is the amplitude plot of the output speech signal. Here we can observe that the frequency of the output signal is higher than that of the input speech signal. We can closely observe the first and the last graphs in the figure to exactly make out the change in the frequency. The frequency has increased as a result of overlapping of windows.

CONCLUSION

By the end of paper phase-I, the recorded signal is successfully transformed by analysing the pitch frequency of the low frequency input signal and manipulating the acoustic features to get high frequency output signal. The work for the solution for one of major unrecognised problems in the society of half way done. In order to implement the paper in real-time will need an input port. The input port has to be connected with the help of microcontroller. The future work for the phase-II of this was to implement the work on hardware device using Raspberry pi microcontroller. This goal is achieved in this phase-II of the paper. Through the USB port of the Raspberry pi 4 model B microcontroller microphone is connected and the Bluetooth speaker is connected wireless to the microcontroller. the input signal is obtained from the microphone in real time and the voice transformation is done on the raspberry pi module by running the MATLAB code. The real-time output is obtained on the Bluetooth speaker. The aim of the paper is achieved by converting the pitch and frequency of the masculine transgender voice to feminine voice.

REFERENCES

- [1] Mark Tse, “**Voice Transformation**”, Columbia University, EE6820 Speech and Audio Processing Paper Report, Spring, 2003
- [2] Liliana, Resmana Lim, Elizabeth Kwan, “**Voice Conversion Application (VOCAL)**”, Informatics Department, Industrial Technology aculty Petra Christian University Surabaya, 2011 International Conference on Uncertainty Reasoning and Knowledge Engineering
- [3] Hirokazu Kameoka, Wen-Chin Huang, Kou Tanaka, Takuhiro Kaneko, Nobukatsu Hojo, and Tomoki Toda, “**Many-to-Many Voice Transformer Network**”, IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 29, 2021
- [4] Allam Mousa, “**Voice Conversion Using Pitch Shifting Algorithm By Time Stretching with PSOLA And Re-sampling**”, Journal of ELECTRICAL ENGINEERING, VOL. 61, NO. 1
- [5] Ali Akbar Shah, Zulfiqar Ali Zaidi, Dr. Bhawani Shankar Chowdhry, Dr. Jawaid Daudpoto, “**Real time ace Detection/Monitor using Raspberry pi and MATLAB**” IEEE 10th International Conference on Application of Information and Communication Technologies (AICT), 2016
- [6] K. Horak and L. Zalud, “**Image Processing on Raspberry Pi or Mobile Robotics**”, International Journal of Signal Processing Systems Vol. 4, No. 6, December 2016
- [7] <https://in.mathworks.com/products/matlab.html>
- [8] <https://www.youtube.com>
- [9] <https://in.mathworks.com/products/matlab.html>
- [10] <https://en.wikipedia.org/wiki/>