International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET

Gender and Age Recognition Using Speech Signal

Takale Rutuja[1], Shendge Vrushali[2] , Pachpute Snehal[3], Dangat Prajkta[4] , Prof. Divekar S. N[5]
Student[1-4], Department of E&TC, HSBPVT's GOI, College of Engineering, Kashti, MH, India
Head of Department[5], Department of E&TC, HSBPVT's GOI, College of Engineering Kashti, MH, India

Article Info
Volume 9, Issue 3

Page Number : 237-242

Publication Issue :
May-June-2022

Article History
Accepted : 10 May 2022
Published: 24 May 2022

ABSTRACT

For a variety of reasons, automatic age and gender detection in voice applications is critical. One of the reasons is that it has the potential to improve human-machine interaction. The telephone it can also aid in the identification of suspects in criminal situations, or at the very least reduce the number of suspects. A list of suspects other applications of this technique include the modification of waiting queue music. Where, depending on the person's age and gender, a different sort of music might be played. And also using this age and gender recognition system .For a variety of reasons, automatic age and gender detection in voice applications is critical. One of the reasons is that it has the potential to improve human-machine interaction. The telephone it can also aid in the identification of suspects in criminal situations, or at the very least reduce the number of suspects. A list of suspects other applications of this technique include the modification of waiting queue music. Where, depending on the person's age and gender, a different sort of music might be played. And also using this age and gender recognition system.

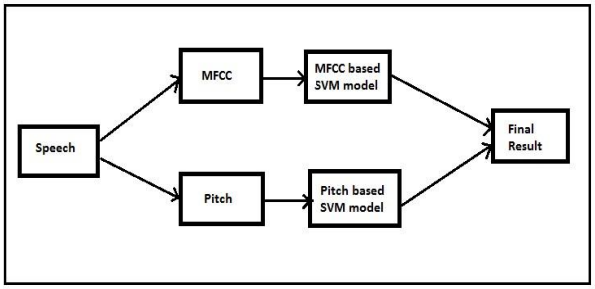
Keywords : Gender Recognition, Sentiment Analysis, Age Detection.

I. INTRODUCTION

The technology has improved significantly over the last decade. With the improvements in new algorithms, big data technologies and data storage methods, it is continuing to improve more. Parallel to these technological improvements, speech recognition systems have improved significantly. Now we are able to talk to our phones to get directions, to ask for some information or to send a text message. Age and gender recognition for speech applications has many practical applications and it can be useful in many applications such as human computer interaction or information retrieval. It can also improve the intelligibility of systems and can be helpful in speaker recognition and surveillance systems. Some researchers tried different speech features and different classifiers in age and gender recognition which can be found under overview of the past history. Researchers have recently used SDC instead of MFCC as a speech feature and so in this research we wanted to explore the viability of this feature for age and gender recognition systems this paper was used[1] ELSDSR read speech corpus is used in this work which is a good database for age and gender recognition. In this work, we used one of the most popular machine learning algorithms that supports Vector Machines (SVM) and tried to predict the age and gender of the speaker. This technique

enables speech recognition systems to personalize the ads according to the person's age and gender and also it can have some uses in criminal cases since most of the proofs are of telephone speeches. Also it can be used for processing waiting queue music for different genders and age groups. Not all people appreciate the same type of music. Older people might like slow music whereas younger people might like rock or metal music. Another usage of this system can be to try to understand age and gender distribution of a population in an experimental study which gives more details about the experiment. Age and gender recognition system consists of two parts. The first part is called pre-processing and feature extraction. The second part is called classification. In the first part, speech is pre-processed using Digital Signal Processing (DSP) techniques and then some of useful features such as MFCC and pitch information are extracted from the speech.

II. BLOCK DIAGRAM



Block Diagram for the Final Fused Age and Gender Recognition System

For getting the other speech feature pitch. The algorithm uses sub harmonic to harmonic ratio to get the pitch information. Also the frequency range was set between 100 Hz and 300 Hz for more reliable pitch estimation. Each frame size was set to 25 ms. after getting the pitch information for all frames in the training set, the mean value for each training set was taken as the pitch of the training example. Model Training Age and gender recognition is a multi-class classification task. After using the previous pre-processing and SDC feature extraction steps, an SVM was trained for both genders and age groups. Nonlinear RBF kernel was used in this test. The database contained 4 labels. These labels included young adult male, young adult female whose age ranges between 20 and 40 years and also middle age male and middle aged female whose age range between 40 and 65 years old. For SVM training, the algorithm described in this paper was used [2]. The parameters of SVM was first selected manually and second time they were selected with cross validation after training on a balanced sub set. For pitch calculation, again the same preprocessing techniques were applied to the training set. After that steps, pitch extraction algorithm which uses harmonic to sub harmonic ratio was executed on the each training sample this paper was used [5].The frame window was set to 25 ms. and after getting the pitch value for each frame in all the training examples, the mean value of each speech example was taken as one feature vector in the training set. To make the model simpler, a threshold value was selected to be 200 Hz. The values below 200 Hz was considered as male speech and the frequencies above 200 Hz was considered as female speech. To use pitch and MFCC score together, they were first scaled to the same dimension and later was done a weighted sum to get the final result. The final fused age and gender recognition system can be seen in Figure.

III. FLOW CHART

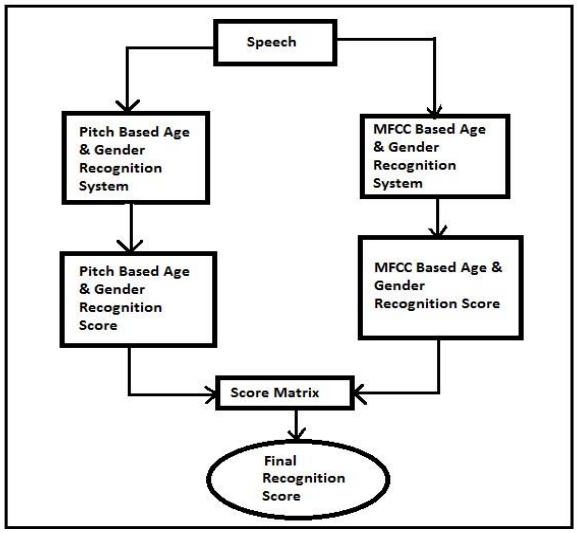
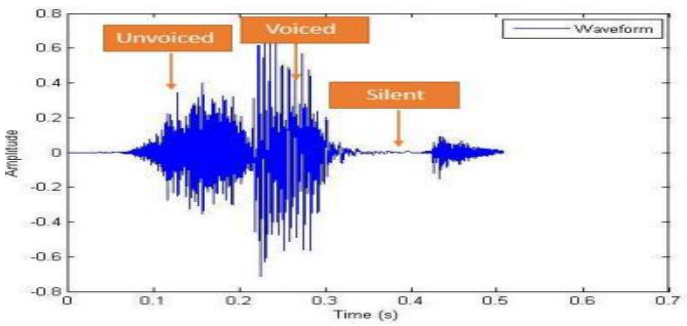


Figure shows a flow chart of our proposed fused age and gender recognition system for speech applications. As it can be seen from the figure, our system consists of two main components. In the first component, age and gender recognition is done based on pitch information of speaker. In the second component, age and gender recognition is done using MFCC which performed better than SDC on ELSDSR corpus. Scores from both pitch based model and MFCC based model are fed into a score matrix and the score matrix generates a final recognition score. Described in this paper was used [3] [4] .

IV. SYSTEM DEVELOPMENT

Speech Processing Background

Speech is a natural way of communication between human beings. Speech is produced by the vocal folds vibrations, movement of articulators and also with breathing of air from the lungs. Sounds created using lips, tongue and mouth positions are based on some rules. Speech signals created by humans are analog in nature. So in order for computers to process speech information, first, the speech must be converted from analog to digital signal. Also, speech signals can be represented in time domain or in frequency domain. While representing speech signals in time domain, on the x-axis there is time and on the y-axis there is amplitude. That does not tell much about the frequency content of the speech. So a better representation would be the frequency domain representation of speech. In this domain, on the x-axis there is frequency and on the y-axis there is magnitude in dB Also, speech signals in time domain can be classified into three categories as voiced, unvoiced and silent speech as shown in Figure 5. Voiced sounds are periodic in nature and have higher energy than unvoiced sounds which are aperiodic and noise-like. The silence is when there is no speech and may have energy level related to the background noise.



Time Domain Classification of the Speech Utterance 'Sit'

In terms of digital modelling of speech production, there are two factors that come into this equation. The first one is the excitation source which can be periodic or aperiodic. When the vocal fold vibration is modulated by the vocal tract filter it generates a periodic voiced sounds and similarly turbulent air flow from the lungs is modulated by the vocal tract it generates aperiodic noise-like unvoiced sounds .Humans can produce speech starting from 50 Hz and usually this range is between 50 Hz to 3400 Hz. But most of the time, the majority of the energy is between 300 Hz and 3000 Hz. Human ear, on the other hand, can hear frequencies between 20 Hz to 20 kHz. Frequencies above 20 kHz is considered as ultrasonic sounds and human ear is not capable of hearing those frequencies. Also, frequencies below 20 Hz is called infrasonic and humans don't hear it. In time domain of speech signals, it can be seen that voiced sounds have a repeating periodic pattern. Each of these identifiable patterns is called a cycle. The duration of a cycle is called the pitch period and the fundamental frequency ($F0$) or pitch frequency is the inverse of pitch period ($T0$). Fundamental frequency shows how high or high low, a person's voice sounds. It is the frequency of his or her vocal cord vibration. Adult males typically have a fundamental frequency between 85 Hz to 155 Hz. Adult females, on the other hand have higher fundamental frequencies. An adult female fundamental frequency is in the range of 165 Hz to 255 Hz. Infants have much higher fundamental

frequencies when they speak. It is generally between 250 Hz to 650 Hz. A ten year old boy or girl has a fundamental frequency of 400 Hz. When a person speaks, his/her fundamental frequency changes because of the structure of the language such as intonation and rhythm. So it is not easy to say that there is just one fundamental frequency of a person. However, when the person speaks in a natural voice, it is considered as his/her fundamental frequency.

V. SPEECH SIGNAL CHARACTERISTICS

When humans speak, the speech signal produced by the vocal tract is an analogue signal. As mentioned earlier, the data on the computers are stored digitally. So when working and processing speech data on computer, first thing, the speech data needs to be converted into digital signal. Also the speech data needs to be sampled at a high rate in order not to lose important information of speech. According to Nyquist–Shannon sampling theorem, the digital sampling frequency must be at least two times bigger than the highest frequency in the analog signal. 5.2.1 Speech Framing and Windowing Speech signal is a time-varying signal. It is stationary and changes over time. So in order for speech to be processed, it must be divided into non-stationary frames. The general size of speech frames varies between 10ms to 40ms where speech is said to be not changing. Figure shows one frame of the vowel /I/ from the speech utterance 'Sit'.

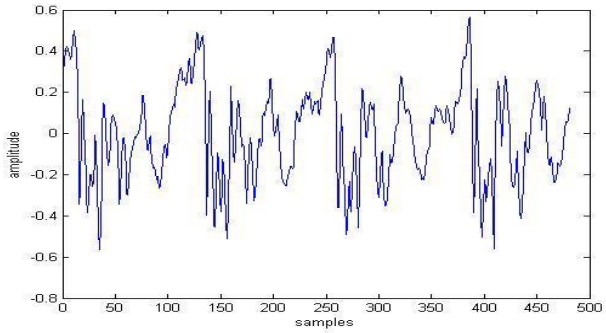
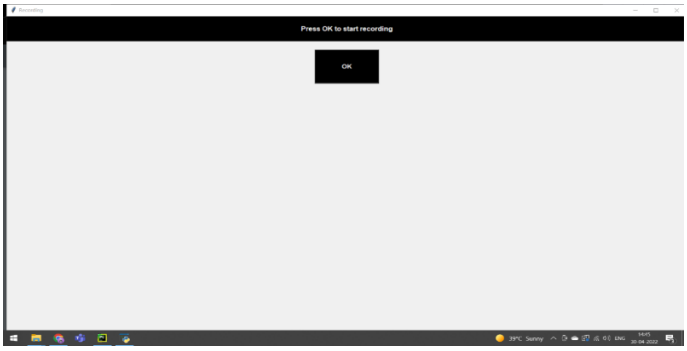


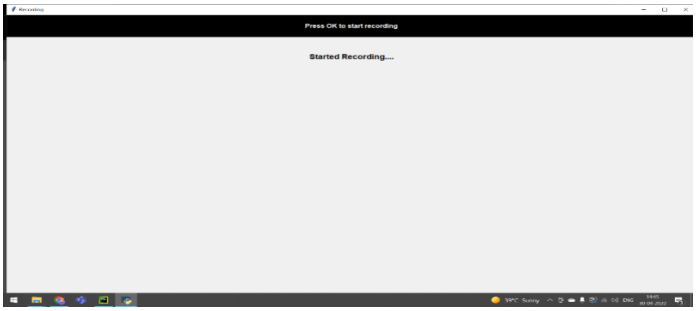
Fig One Frame (30ms) of /I/ Sound from the Speech Utterance 'Sit'

VI. RESULTS

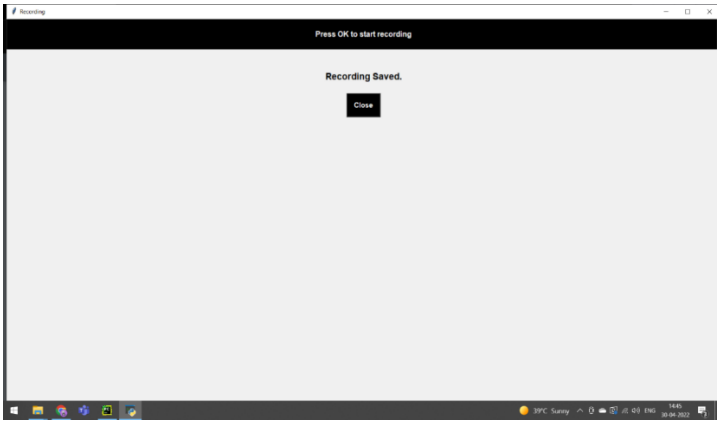
In the beginning only Record voice option is available as user needs to record the voice first to check gender.



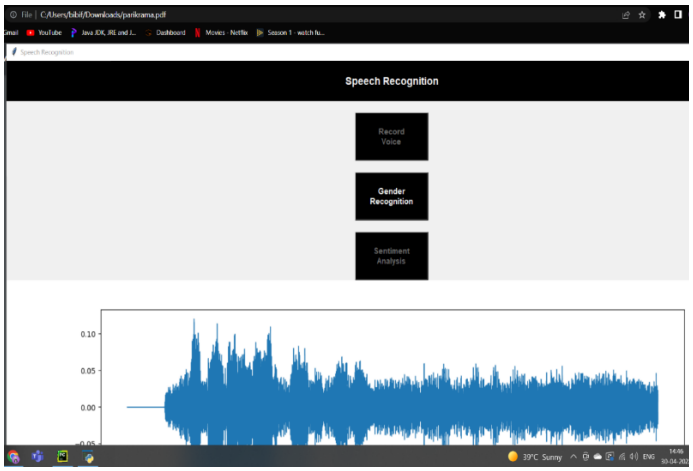
After clicking on record voice it ask to press ok to record the voice. And then after 5 seconds the is getting started to record which is shown in below figure.



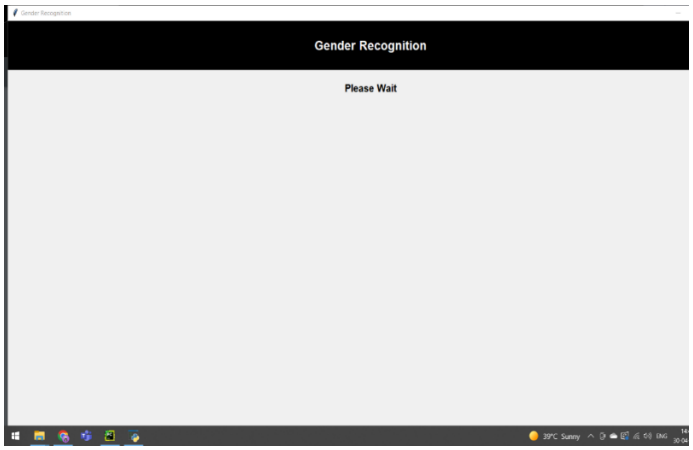
This snapshot shows the voice is being recorded.



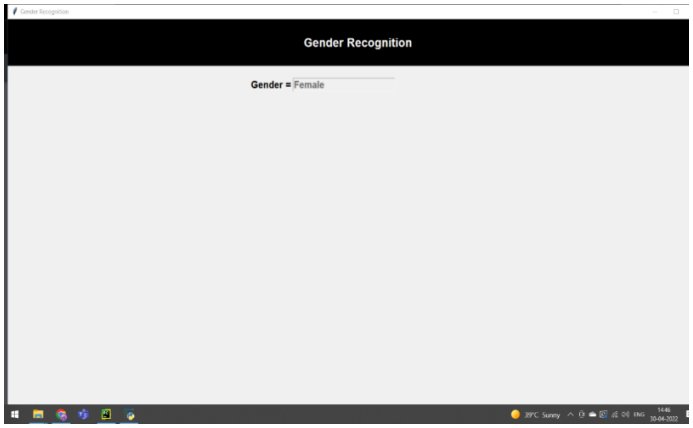
When user stop speaking then the voice detection process stops and it shows a message "Recording Saved"



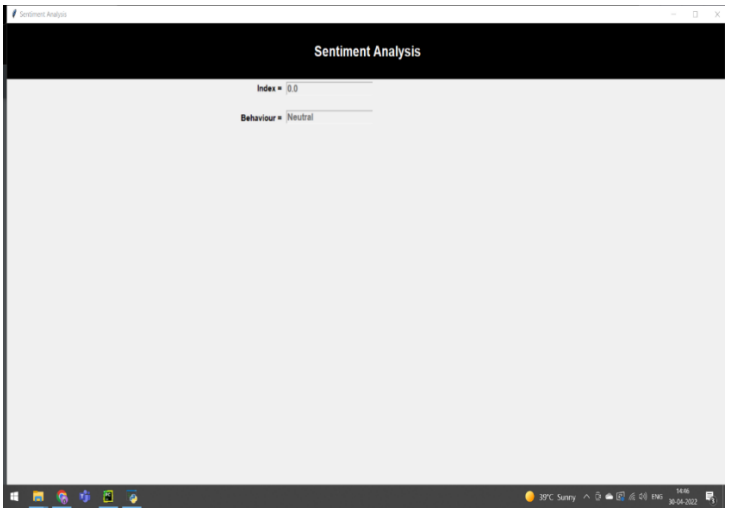
After the recording we can check gender by clicking on Gender Recognition option.



After clicking on Gender Recognition it shows a message “Please Wait”



After that it recognize the gender in which the voice is recorded.



We can check the behavior also by choosing the option of Sentiment Analysis

VII. CONCLUSION

In this paper the major objective was to create a speech-based gender recognition system. One of the most critical aspects of building a gender recognition system is feature selection. According to a review of past studies, the power spectrum has the best classification rate among the various parameters. As a result, the power spectrum has been chosen as the categorization characteristic. The statistical analysis and threshold approach, among the other techniques, is straightforward to compute and delivers excellent results. This is why, in order to boost performance, this approach was chosen for pattern comparison in the recognition process. The average accuracy of recognition is 80%. The recognition rate drops as the number of speakers increases, according to the testing results.

Speech recognition is an incredible human ability, especially when you consider that a regular conversation involves the recognition of 10 to 15 phonemes every second. It should come as no surprise that developing machine (computer) identification systems has proven tough. Despite these issues, a growing number of systems are becoming available that have some success, generally by focusing on one or two specific areas of speech recognition.

Speech synthesis systems, on the other hand, have been around for quite some time. These technologies are increasingly a frequent part of our lives, despite their limited capabilities and lack of the "natural" character of human speech.

VIII. REFERENCES

[1]. L. Feng, “Speaker Recognition, Informatics and Mathematical Modelling”, Technical University of Denmark, 2004M. Sahidullahl, “Mfcc to SDC Matlab Library”, 2011

[2]. C. C. Chang, and C. J. Lin, "LIBSVM: a library for support vector machines." ACM Transactions on Intelligent Systems and Technology, 2011

[3]. T. Bocklet, G. Stemmer, V. Zeissler, and E. Nöth, "Age and gender recognition based on multiple systems-early vs. late fusion" INTERSPEECH, 201028 65

[4]. M. Li, K. Han, and S. Narayanan, "Automatic speaker age and gender recognition using acoustic and prosodic level information fusion," Computer speech and language, Vol. 27, No. 1, pp. 151-167, Jan. 2013

[5]. X. Sun, "A pitch determination algorithm based on sub harmonic-to-harmonic ratio", 2000

Cite this article as :

Takale Rutuja, Shendge Vrushali, Pachpute Snehal, Dangat Prajkta, Prof. Divekar S. N, "Gender and Age Recognition Using Speech Signal", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 9 Issue 3, pp. 237-242, May-June 2022.
Journal URL : https://ijsrset.com/IJSRSET2293108