

Fuzzy Logic Based Voice Recognition as per Their Gender and Age Group

J. Rajaram

Abstract-Human voice is the very important field in digital speech processing. In the coming technologies every system will be based on the human voice. The system will be locked and unlocked by human voice. In this research, we classify the voice according to their age groups. We can recognize the voice by calculating many various parameters such as pitch, power amplitude and THD and many more. There are many approaches through which we can recognize the gender's voice such as Hidden Markov Model [HMM], Dynamic Time Warping [DTW] etc. This paper is basically depending on the voice perception system which helps in to differentiate the voice between the gender's voice and also differentiate the voice depends on age gp to 12yrs children voice with the use of SVM and Gaussian membership function.

Keywords—voice recognition system, fuzzy logic, genetic algorithm, age based voice classification, Support Vector Machine(SVM), Gaussian Mixture Model(GMM)

I. INTRODUCTION

Speech recognition is the technique to convert the input speech signal (speech signal) to the digital signal on which we apply some techniques and methodologies to find the voice according to their age group. Pitch is the main factor in the speech recognition because every age group there is different pitch value it means lower age group pitch is too high and senior age group pitch is too slow. There are several applications in which we use voice as a security such as Defensive Purpose, Bank Security, Investigative actions etc. In today's life every person have their own gadget but there are some people who are unable to lock properly and anyone can misuse it but through voice we can lock the gadget and only the owner can unlock the mobile phones, laptops etc. So, the voice recognition is very important and in this paper, we define that there are many methods to understand the difference between the male and female voice, to recognize the voice depending on their age group till 12 years child.

II. VOICE RECOGNITION

Voice recognition is the important part of human conversation. The human voice can be used for security purpose, it means we can lock or unlock the system by recording voice and another person can't unlock the system. Voice recognition is the type of software invention which helps in communicating the human with the mobile devices and many other electronic devices. In an early generation people do know how to read they can't access the electronic devices properly so through voice recognition they can access the devices easily.

The researcher records many voices of different age groups of different values. Some voices contain utterance, silence and many others things. Firstly we will eliminate the silence from the voice, then we collect the features of the voice and then we calculate the values of the parameters. Depending on the values we recognize the voice it means the particular voice belongs to which age group.

Till now the work is done only on the gender recognition system in which they classify the difference between gender's voice. The classification of male and female voice is already implemented using genetic algorithm and it also implemented in fuzzy logic toolbox and also with the use of the neural network tool.

But the classification of voice recognition system depends on age is new because it is not implemented yet now. The age based recognition system is very beneficial for a human being because it defines the difference between the voice of different age groups. We mainly calculate three parameters pitch, power amplitude and total harmonic distortion by using its formula. After calculating the values we get the range of each voice signal, then we create the fuzzy rules according to its values. Before making the fuzzy rules we have to calculate the range of each parameter, then fuzzy rules are applied to classify the voice according to its age group.

There is a procedure to classify the voice according to its age group:-

1. We have to set the age groups on which we decide the groups are to be taken like in this research, we take 5 age groups which are defined below
 - a. Small children (0-5)
 - b. Children (06-14 years)
 - c. Youth (15-24 years)
 - d. Adults (25-64 years)
 - e. Seniors (65-80 years)
 - f. Very seniors(81 years and over)
2. After this we record the samples of each age group.
3. After record the samples we convert the voice samples into signals.
4. With the help of the signals we calculate the pitch value, power amplitude
5. Then we make the fuzzy rules.
6. After applying the fuzzy rules we can easily classify the voice of each age group according to their pitch value, power amplitude value and total harmonic distortion values.

Revised Version Manuscript Received on July 18, 2019.

Dr. J. Rajaram, Professor Dept of Computer Science & Engineering, Brilliant Grammar School Educational Society's Group of Institutions-Hyderabad, Telangana, India.

III. FUZZY LOGIC

Fuzzy logic is also called as fuzzy set theory, which was established in 1965 by Lofti Zadeh to combat the problems mathematically. It is completely based on “de-gree of truth” rather than the values which is given in true or false. The application of fuzzy logic based consumer based devices such as cameras, washing machines and medical instrumentation etc. Fuzzy logic has many different meanings. In the basic sense, this is a logical system that represents the symbol of multivalued logic. On the large scale, fuzzy logic (FL) is usually synonymous with the fuzzy set, which belongs to classes of the objects in which boundaries are not clear and it includes a membership function which is also called a degree of matter [12]. Fuzzy sets are the superset of classical sets in which each element is affiliated with the real number which represents the degree of membership of the elements which are present in that particular set. In the fuzzy logic we have some inputs which we apply on some inference system like FIS and gets the output results.

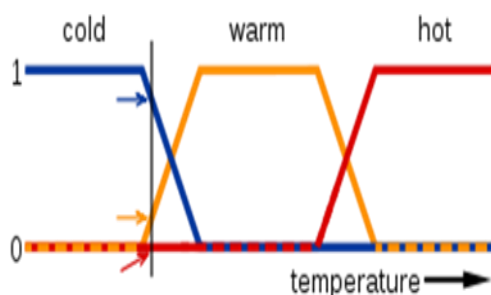


Fig1: Fuzzy Logic[12]

The basic concept which comes under the fuzzy logic is that of the linguistic variable it means the variables whose values are words not the number it doesn't give the answer in 0 and 1. In the fig 1 it shows that the values come under the range 0 and 1. It is clearly shown that the temperature is not exactly identified. There are many possibilities which defines the temperature like:

1. Very cold
2. Cold
3. Warm
4. Hot
5. Very hot

This shows that the fuzzy logic has linguistic variables which don't give the exact value like 0 or 1.

Fuzzy logic is a multivalued logic, which means we can apply the fuzzy rules on more than one variable and we used a fuzzy inference system to get the output. There are mainly two types of methods which comes under FIS:-

1. Mamdani's fuzzy inference method
2. Sugeno's type system

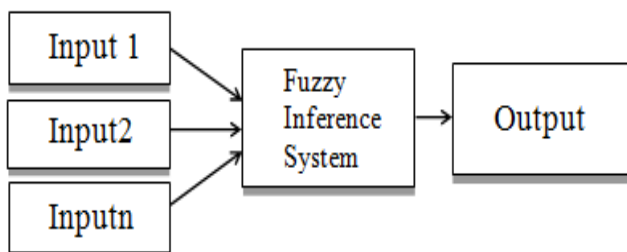


Fig 2: Fuzzy Inference System

Generally, we used Mamdani's fuzzy inference system because it is seen that this is mostly used in fuzzy methodology and it was the first control system which was built by using the fuzzy set theory. Mamdani's type fuzzy inference was established in 1975 by Ebrahim Mamdani [14] as he tried to control a steam engine by orchestrate the set of linguistic fuzzy rules from the qualified human operators. This type of FIS is defined for the toolbox, which assumes that the output of the mf are to be the fuzzy sets and after this process of gathering the membership function that needs defuzzification process to get the output.

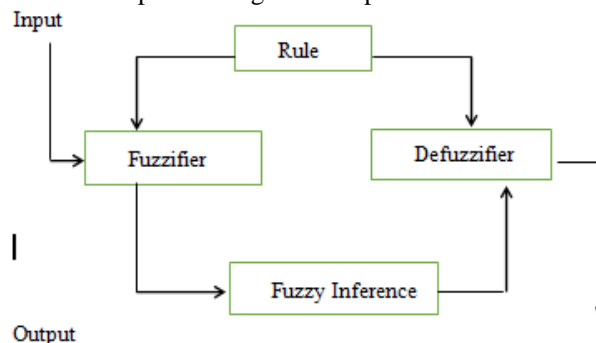


Fig 3: Architecture of Fuzzy Logic

In the fig [3] it describes the whole scenario of fuzzy logic, means how the input values are converted into crisp value called fuzzification, then we applied some fuzzy rules on the crisp values of the input. After that we use fuzzy inference system and we get some output variables and then we defuzzify the system or output variables to get the output results.

IV. METHODS

There are many methods which help in differentiating the gender's voice. We also classify the different age's voice and we recognize the voice according to their age group it means we identify which voice belongs to which age particularly. If any fake person wants to enter into the system, then the system should recognize that he/she is not the owner of the system and it shows an error. There are many methods to classify the voice like pitch, its frequency, power amplitude, total harmonic distortion, and to collect the feature we use MFCC (Mel-frequency Cepstral coefficients) is the technique which helps in collecting the feature of the speech and many more.

4.1 Features Extraction-

This is the initial step to recognize the speech of any person. For feature extraction we mainly use MFCC which helps in getting the exact output.



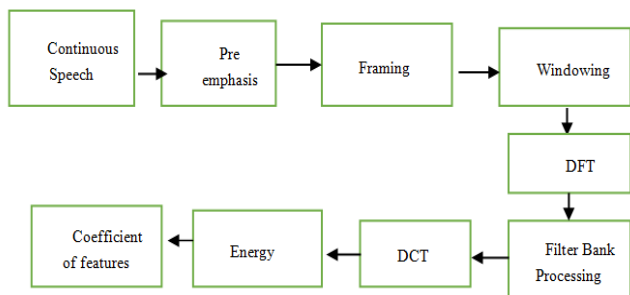


Fig 4: Feature Extraction [17]

In the fig 4 it describes the whole process to extract the features of the voice[17].

1. Continuous Speech: In the first step we collect the continuous speech with the use of a microphone.

2. Pre-emphasis: To collect higher frequencies we preprocessed the signal through the filter banks which emphasizes the frequencies.

3. Framing: In this step the segments of speech are created and these segments are getting into small frames within the range of 20 to 40 ms from the conversion of analog to digital also known as ADC.

4. Windowing: In this process, we simply reduce the distortion of the signal. The window is simplicity used when we retrieve the N sample frames from the signal. It helps in reducing the distortion, but it increases the single shape alteration, which is mostly used in window shape is hamming. And this hamming window is used for considering the next step of feature extraction process.

5. DFT- DFT is known as Discrete Fourier Transform which is the standard method of spectral analysis, which rely on the Fourier Transform which helps in converting into the frequency domain of each N samples. It helps in reducing the computational complexity to order lg(N).

6. Filter Bank Processing: Generally, the spectral features are obtained in the next process of the filter bank, which properly integrates the spectrum at defined frequency range.

7. DCT: It is also known as Discrete Cosine Transform which is used to generate the highly uncorrelated feature, DCT used to achieve Mel-cepstrum coefficients through which we are going to select 13 coefficients for a speech recognition system.

8. Energy: The energy of the voice signal such as the slope of formant must its transitions, it means it is necessary to add the features according to change in cepstral features over time. And this cepstral coefficients are joined to energy co-efficients using the algorithm of the energy of the frames.

4.2 MFCCs-

The Mel-Frequency Cepstrum is the technique which we used to extract the features of the voice and this is based on frequencies (which are lower than 1000 Hz) with the well-known fluctuations of human ear's critic bandwidth. The motive of the MFCC is to describe about the behavior of the human voice. The detailed process of MFCC is-

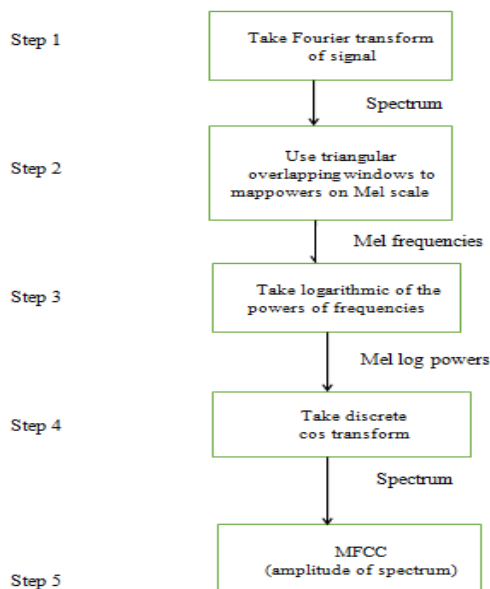


Fig 5: MFCCs Process [16]

In the above figure it describes the detailed progress of the MFCC of the voice it means we easily find out how the human ear behaves. In the recognition of the speech either on the basis of the gender recognition or the age base voice recognition. Initially, we should know about the features of the voice and to extract the features of the voice we use MFCC spectrum. The Mel-frequency scale of linear frequency spacing is below 1000 Hz and log spacing above 1000 Hz. The accuracy of the voice recognition system is very high it means the performance rate of the MFCC is high. Finally, MFCCs, $C[n]$, were calculated by [1][17]

$$c[n] = \sum_{m=0}^{M-1} \ln[FBE(m)] \cos\left(\frac{\pi n(m+1/2)}{M}\right) \quad n = 0, 1, \dots, M \quad (1)$$

where, M is the number of Mel bands calculated by:

$$FBE[m] = \sum_{k=0}^{N-1} |X[k] 2Hm[k]| \quad m = 0, 1, \dots, M \quad (2)$$

Where $X[k]$ is FFT of input signal ($x[n]$). $Hm[k]$ is the triangular weighting windows and N represents the number of FFT points. The formula to calculate the Mels for a given frequency f is calculated by:

$$f[\text{Mel}] = 2595 \log_{10}(1 + f[\text{Hz}] / 700) \quad (3)$$

4.3 Pitch-

Pitch is the most important feature to differentiate the voice either between gender's voice or age based voice recognition system. Pitch can be calculated by the pitch detection algorithm. With the help of the calculated values from pitch detection algorithm, it is proved that the pitch of the female voice is high as compared to male voices. According to the researcher's the pitch of the human's voice is decided by the resonant frequency of the vocal folds. The values of the pitches are very different from each other the pitch of the small children is very high from the adult male and female pitch value. The frequency of adult male is about 125Hz, in the adult female the frequency is about 210, and in the frequency of children's voice is more over 300 Hz. The vocal folds are vibrating when they are closed to obstruct the airflow through the glottis and this principle



is known as Bernoulli's principle[17].

4.4 Short-term Energy-

This is the method which helps in finding the voiced and unvoiced part in the speech signal. The amplitude of the human voice signal varies with the time. The amplitude of the unvoiced part of a human voice is lower and the amplitude value of the voiced part is higher. This can be calculated by this equation [4]

$$E = \sum_{m=-\infty}^{\infty} [x(m)w(n-m)]^2$$

4.5 Total Harmonic Distortion-

Total harmonic distortion or THD is a quantification of the harmonic distortion existed in a signal and is described as the ratio of the sum of the powers of all harmonic components to the power of the fundamental frequency. From the testing results it is seen that the total harmonic distortion of the female voice is higher than the male voice.

4.6 Gaussian Mixture Model

A Gaussian mixture model is a probabilistic model that speculates all the data points are originated from a mixture of a finite number of Gaussian distributions without known parameters. A Gaussian mixture object implements the expectation-maximization (EM) algorithm for fitting Gaussian mixture models. A GaussianMixture.fit method is provided that learns a Gaussian Mixture Model from trained data.

4.7 Support Vector Machine

SVM is the most energetic machine learning technique which is used for data classification and pattern classification. The authors applied the SVM to the acoustic model of voice recognition system which are based on the Linear Predictive Coding and MFCC features for the different dataset [16]. SVM comes under the supervised machine learning technique that helps us in analyzing the data which we can use for classification and also for regression analysis.

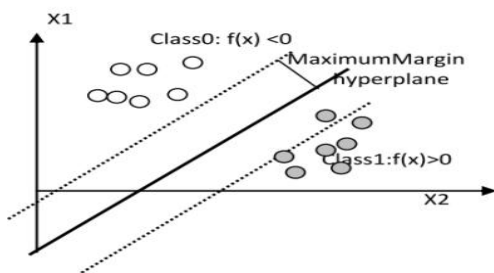


Fig6. Support Vector Machine [8]

Where, t_i represents the ideal outputs, $i > 0$ and the vector X_i is support vector and formed using the training set by an optimization process.

V RESULTS

The aim of this research is to classify the different voice according to the male and female voice and as per their age group. There are many techniques which helps in extracting the features of the voice like PLC, MFCC etc. MFCC is the best method for extracting the features of the voice. There are many parameters which we can use for the voice recognition

but in this review, it is shown that the voice can be classify on three parameters such as power amplitude, pitch and total harmonic distortion. Till now the work is done at max 50yrs old person's voice. The best way is to extract the features on the voice dataset, record the voice of different age group with the microphone and use either machine learning technique or fuzzy logic and other different methodologies. In this research it is shown that the speech voice recognition can be done with the help of SVM and Gaussian mixture model.

VI CONCLUSION

In this research it shows that there are many approaches which help in to recog-nize the voice on the basis of their age and differentiate the voice between male and female. There are many parameters which are used to differentiate the voice between male and female such as pitch and much more. Pitch is calculated by a pitch detection algorithm and it is proved that the pitch of the females is higher in comparison to male only in the case of voice based gender recognition system. In age based voice recognition system the pitch is different in every age group, it means the pitch of small children is very high and according to the increment in their age the pitch gets slow down but generally there is not so much difference in the female voice. In this research it is shown that the voice of the particular person belongs to which age group.

REFERENCES

1. Iman Esmailia, Nader Jafarnia Dabanloosa, Mansour Valib "Automatic classification of speech dysfluencies in continuous speech based on similarity measures and morphological image processing tools", Elsevier Ltd., 2016.
2. Lokhande, Navnath S Nehe, Pratap S Vikhe, "Voice Activity Detection Algorithm for Speech Recognition Applications", International Conference in Computational Intelligence, 2011.
3. Arjuwan M. Abduljawad Al-Jawadi, "Speech Recognition And Retrieving using Fuzzy Logic System", Technical College, Foundation Of Technical Engineering, Mosul, Iraq, 2009.
4. Bachu R.G., Kopparthi.S., Adapa B., Barkhana B.D., "Seperation Of Voiced and Unvoiced using Zero crossing rate and Energy of the Speech Signal", University Of BridgePort.
5. Atif Khan, Vikas Kumar, Santosh Kumar, "Speech Based Gender Identification Using FuzzyLogic", International Journal of Innovative Research in Science, Engineering and Technology, 2017.
6. Gayathri S, Mugundhan B, "Identification Of Age Using Voice Recognition", International Journal Of Advances in Electronics and Computer Science, ISSN: 2393-2835, Vol. 4, Issue-7, 2017.
7. Shivaji J Chaudhari, Ramesh M Kagalkar, "Methodology for Gender Identification, Classification and Recognition of Human Age", International Journal of Computer Applications (0975-8887), National Conference on Advances in Computing, 2015
8. Ahuja Pooja, Vyas JM, "A Development Overview Of Voice as a Steadfast Identification Technique", Journal of Forensic Research, ISSN: 2157-7145, Vol 6, Issue 3, 2015.
9. Syed Mostafa Mmirhassani, Alireza Zourmand, and Hua-Nong Ting, "Age estimation based on children's voice: A fuzzy-based decision fusion strategy", The Scientific World Journal, Vol 2014, Article ID 534064.
10. Venkatesan, C., P. Karthigaikumar, and S. Satheeskumaran. "Mobile cloud computing for ECG telemonitoring and real-time coronary heart disease risk detection." Biomedical Signal Processing and Control 44 (2018): 138-145.
11. Yong-Qian Ying, Peng-Yung Woo, "Speech



- Recognition Using Fuzzy Logic", IEEE, 1999
12. Sahidullah, Md.; Saha, Goutam (May 2012), "Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition". *Speech Communication* 54 (4): 543-565. doi:10.1016/j.specom.2011.11.004.
 13. X. Huang, A. Acero, H.W. Hon, *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*, Prentice Hall, New Jersey, 2000, October.
 14. https://sciencedaily.com/terms/vocal_folds.htm
 15. https://en.wikipedia.org/wiki/Vocal_cords Kamil Aida-zade, Anar Xocayev, Samir Rustamov, "Speech recognition using Support Vector Machine", IEEE, 2016
 16. Shivaji J. Chaudhari, Ramesh M. Kagalkar, "Automatic Speaker Age Estimation and Gender Dependent Emotion Recognition", *International Journal of Computer Applications*,(0975 - 8887), Volume 117 - No. 17, May 2015.