International Conference on Advancements in Computing Technologies - ICACT 2018 ISSN: 2454-4248
Volume: 4 Issue: 2 130 – 133

Automatic Gender Recognition using ANFIS

Radha N[1], Malathy E. M[1], Jeya Aiswarya V[2]

Assistant Professor, Department of Information Technology,SSN College of Engineering, Kalavakkam, Chennai,India[1]

U.G. Student, Department of Information Technology, SSN College of Engineering, Kalavakkam, Chennai, India[3]

Abstract: Results on classifying a speaker on the basis of gender by processing speech and analyzing the voice samples are presented in this work. First, the speech samples are collected from different speakers. With the help of speech signal processing, a speech signal is analyzed first, followed by extraction of features from the speech signal. Feature extraction is performed over the speech by using cepstrum analysis to determine the pitch of a speech signal. The Pitch of various speech signals is given as an input to train ANFIS network. Finally, an automated gender classification is successfully performed by ANFIS, although, the ANFIS has to be trained before the actual classification.

Keywords:*ANFIS-Adaptive Neuro-Fuzzy Inference System, Gender recognition, Cepstrum analysis*

*****

I INTRODUCTION

In modern civilized societies for communication between human speeches is one of the common methods. Different ideas formed in the mind of the speaker are communicated by a speech in the form of words, phrases, and sentences by applying some proper grammatical rules. By considering speech as one of the outcomes of passing a glottal excitation waveform through a time-varying linear filter which can be used to represent a speech signal, so as a speech production model that models the resonant characteristics of the vocal tract [1]. By classifying the speech with voiced, unvoiced and silence (VAS/S)an elementary acoustic segmentation of speech which is essential for speech can be considered. In succession to individual sounds are called phonemes and this technique can almost be identical to the sounds of each letter of the alphabet which makes the composition of human speech. Speech processing is the study of speech signals and the various methods which are used to process them.

In this process various applications such a speech coding, speech synthesis, speech recognition and speaker recognition technologies are the concepts used in speech processing is employed. Among the speech processing application, speech recognition is the most important one in many real-time application. The main purpose of speech recognition is to convert the acoustic signal obtained from a microphone or a telephone to generate a set of words. In order to extract and determine the linguistic information conveyed by a speech wave, we have to employ computers or electronic circuits. This process is performed for several applications such as security device, household appliances, cellular phones ATM machines and computers. Gender classification is one of the humans to machine interaction technique is applied in many fields. For example, it is applied in various applications such as speech recognition, speaker diarization, speaker indexing, annotation and retrieval of multi-media database, synthesis, smart human-computer interaction biometrics social robots etc. and it is a difficult and challenging problem. We can identify physiological differences such as vocal fold thickness or vocal tract length and differences in speaking style of humans as partly the reason gender-based differences in a human speech [2].

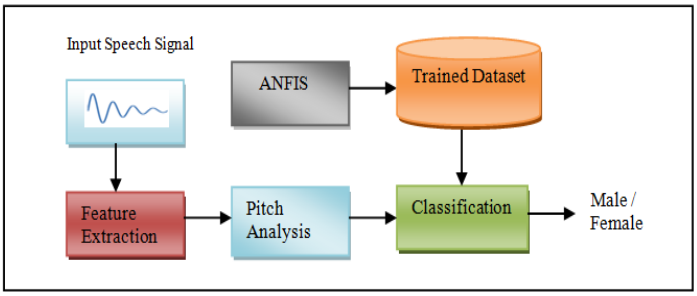

Fig 1. Block diagram of a proposed work

Normally the higher formant frequencies and fundamental Frequency (FO) are higher for female speakers and the FO differences are larger than the formant frequency differences between male and female groups. Female speakers various speech qualities or characteristics like aggressiveness, body size, self-confidence, and assertiveness are related to low FO. The core problem addressed in this research work is automated pitch based gender recognition imbibing into a machine the capability of identifying the gender of a human speaker, that is, to identify whether a human speaker is male or female. Pitch-based gender recognition based on speech signal processing has been developed using MATLAB to automate the classification process. With the help of speech signal processing, a speech signal is analyzed first, followed by extraction of features from the signal using cepstrum method. During speech analysis, the speech signal is processed and classified into voiced/unvoiced/silence segments for enabling the extraction of the pitch. Feature extraction is performed over the classified speech by using cepstrum analysis to determine the pitch of a speech signal. This allows pitch-based gender recognition to be accomplished on the basis of the extracted pitch value using a trained Adaptive Neuro-Fuzzy Inference System (ANFIS). Figure 1 shows a schematic block diagram of the proposed system.

ISSN: 2454-4248
130 – 133

Paper is organized as follows. Existing work relevant to gender recognition is presented in Section II. Section III describes feature extraction module and gender recognition using ANFIS. The flow diagram represents the step of the algorithm. Section IV presents experimental results showing results of the recorded signal. Finally, Section V presents the conclusion.

II RELATED WORK

Some of the recent research works related to speech classification are discussed as follows. In [3] proposed the design of a computer program to model acoustic analysis of voices and speech for determining gender. The model is constructed using 3,168 recorded samples of male and female voices, speech, and utterances. The samples are processed using acoustic analysis and then applied to an artificial intelligence/machine learning algorithm to learn gender-specific traits. The resulting program achieves 89% accuracy on the test set.[4] have presented two different models by using speech processing techniques and algorithms for gender recognition and one of their model is used to produce formant values of the voice sample and the other model to produce pitch value of the voice sample. The gender-biased features and pitch value of a speaker were extracted by employing these two models. The mean of formants and pitch of all the samples of a speaker were calculated by applying a model having loops and counters which generates a mean of first formant and pitch value of the speaker. The speaker is classified between Male and Female by computing Euclidean distance from the mean value of males and females of the generated mean values of first formant and pitch by using the nearest neighbor technique. Using NI Lab VIEW, the algorithm is implemented in real time [7].

In [5] have discussed that the background noise from the noisy environment for example car, bus, babble, factory, helicopter, street noise and more have reduced the performance of speech processing systems like speech coding, speech recognition etc. Thus the classification of noise is necessary to improve the performance of the speech recognition system. The selection of an excellent set of features that can efficiently separate the signals in the feature space was an important process in the design of a signal classification system. Noise classification is a crucial process in order to reduce the consequence of environmental noises on speech processing tasks. They have proposed a fuzzy ARTMAP network and modified fuzzy ARTMAP network to classify the various background noise signals. Moreover, in addition to it their experimental results were compared with both backpropagation networks and Radial Basis Function Network (RBFN) [6].

.

III FEATURE EXTRACTION AND MODELLING

A cepstrum is a result of taking the inverse of the Fourier transform of the logarithm of the spectrum of a signal. The power cepstrum (of a signal) is the squared magnitude of the Fourier transform of the logarithm of the squared magnitude of the Fourier transform of a signal. The power cepstrum of signal is given by

$$|F\{\log(|F\{f(t)\}|^2)\}|^2$$

The cepstrum is a common transform used to gain information from a person's speech signal. It can be used to separate the excitation signal (which contains the words and the pitch) and the transfer function (which contains the voice quality). Cepstrum plots of male and female voices show that a female voice has peaks occurring more often in it than in the male cepstrum. This is due to the higher pitch of a female voice [6]. Cepstrum analysis was used as part of this work to extract the pitch from the voiced portion of a person's speech sample. An approach to recognizing the gender of a speaker, that is, whether a speaker is a male speaker or a female speaker, was followed in the form of several steps. Once a speech/voice sample has been recorded, further processing is accomplished according to the steps shown in the block diagram (Fig.1) of pitch-based gender recognition. The next algorithm used for our work is presented as follows.

Algorithm: Pitch-based Gender Recognition using an ANFIS:

1. Input the speech sample.
2. Extract the voiced portion of the selected speech sample for pitch extraction on
the basis of an amplitude threshold.
3. Calculate the pitch from the voiced portion using the cepstrum method.
4. Train an ANFIS network to classify pitch into two fuzzy sets, namely, male
and female.
5. Use the trained ANFIS network for pitch-based gender recognition.
6. Recognition of gender category

The sample datasheet used for training a database is given in Table 1. Estimated pitch value from feature module using cepstrum analysis sample data is given in the second column of the table. The experimental analysis used in our work is discussed in the next section.

Table 1. Sample Data for Training

Sample No	Sample Pitch value from Feature Extraction	Gender Classification Result
1	94.4134	1
2	157.0577	1
3	199.0865	1
4	207.4645	1
....	………	………
33	350.365	2
34	410.256	2
35	380.952	2
36	372.093	2
37	470.588	2

IV EXPERIMENTAL RESULTS

The experiment phase is divided into 3 modules to maintain modularity and reusability. The 3 modules are:

 ISSN: 2454-4248
130 – 133

- Getting speech input from the user and calculating pitch
- Training ANFIS network
- Testing calculated pitch with trained ANFIS network

Getting speech input from the user and calculating pitch: This module allows the user to record or load a speech sample. The user can play the loaded sample with a player option. Once the speech signal is loaded, the signal will be plotted in the plot area. The user has to enter start and end frame number to find the pitch. By clicking the pitch button, the user can calculate the pitch. Once the pitch was clicked, the pitch period and pitch frequency are shown. Second, Training ANFIS network: The Pitch was calculated for various samples which were collected from various speakers. The calculated pitch was recorded in a database. The first column in the sheet represents the pitch and the second column represents the index(one for male and two for female). By clicking Anfis button, this function will be called. The output FIS structure was written to separate .fis file3. Testing calculated pitch with trained ANFIS network: The calculated pitch value was given as an input for this module. This module is called after the second module by clicking the ANFIS button. The input pitch was tested against the trained network and the output either male or female was displayed in the main GUI window. Figure 2 through 4 shows the experimental analysis with processing steps of gender recognition

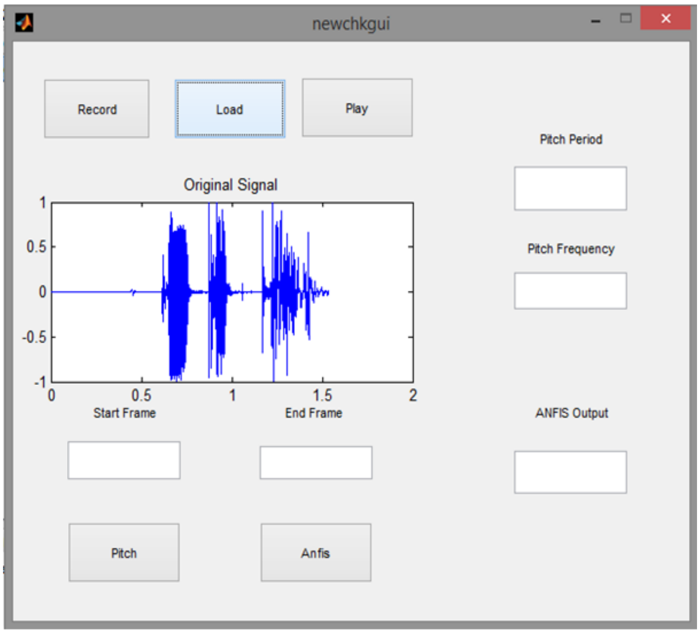

Fig 2. Depicts the original signal

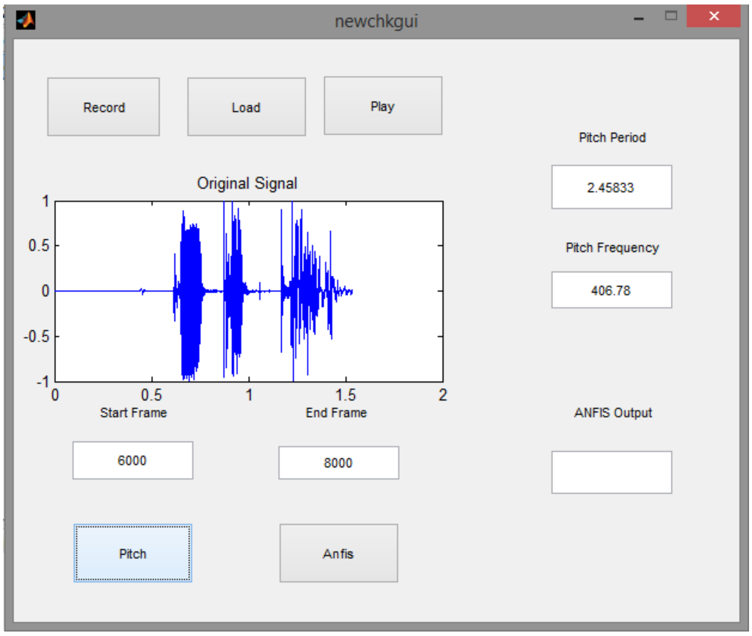

Fig 3. Calculating the pitch by pressing the pitch button

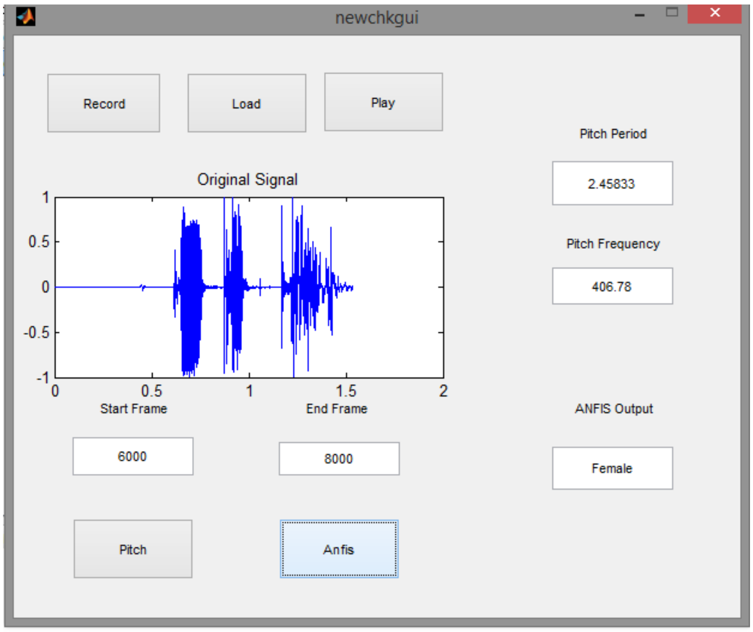

Fig 4. Identifying the gender by pressing the ANFIS

The age group of the speakers ranges between 20-25 and 10 samples per speaker. The language used by the speaker is English. The ANFIS network has correctly identified 29 male samples out of 30 and 28 female samples out of 30.
The table has been formulated below:

Table 2 . Data sets

Age Groups	No. of Samples	Gender	Language used by the speaker	No. of Samples correctly identified
20-25	10(per speaker)	Male	English	29
20-25	10(per samples)	Female	English	28

Table 3 . Accuracy of Gender Recognition

Method	Gender	Recognition accuracy (%)
ANFIS	Male	96.55%
ANFIS	Female	93.33%

The ANFIS has been used to obtain fast and efficient performance of a system. In these two cases, i.e., male and female gender recognition, the system is 96.66% and 93.33% efficient, respectively. The overall efficiency of the system was found to be 94.99 %. This work can be extended for videos and other type of acoustic speech signals[8,9].

V CONCLUSION

In this research work, an automatic gender based classification has been implemented using speech processing of a speech signal. The classification is necessary to calculate the pitch of a speech signal via cepstrum analysis as pitch is present only in the voiced portion of the speech. At the last step, the classification has been performed by a trained ANFIS. The ANFIS has been used to obtain fast and efficient performance of a system. In the two cases, i.e., male and female gender recognition, the system is efficient. In future, it can be extended to next level. The same system can be implemented by using any other algorithm and performance can be analyzed.

REFERENCES

[1]. S. Furui, Digital speech processing, synthesis, and recognition, CRC Press, New York, 2001.

[2]. B. Gold, N. Morgan, Speech and audio signal processing: Processing and perception of speech and music, John Wiley and Sons, Inc, New York, 2000.

[3]. M. Jessen, Speaker classification in forensics and acoustics, Speaker Classification I, LNCS (LNAI), Springer 4343(2007) 180-204.

[4]. C. Müller (Ed.), Speaker classification I: Fundamentals, features, and methods, LNCS, Springer 4343(2007).

[5]. K. Rakesh, S. Dutta, K. Shama, Gender recognition using speech processing techniques in Lab VIEW, International Journal of Advances in Engineering and Technology, 1(2) (2011) 51-63.

[6]. J. Vieira, F.M. Dias, A. Mota, Neuro-fuzzy systems: A survey, 5th WSEAS International Conference on Neural Networks and Applications, Udine, Italy, 25-27 March, 2004.

[7]. K. Wu, D.G. Childers, Gender recognition: Part I, J. Acoust. Soc. Am, 90(4) Part I (1991) 1828-1840.

[8]. Radha, N. 2016,"Video retrieval using speech and text in video", International Conference on Inventive Computation Technologies (ICICT), IEEE, Coimbatore, India.

[9]. N. Radha, A. Shahina, G. Vinoth, and A. N. Khan, "Improving recognition of syallabic units of hindi language using combined features of throat microphone and normal microphone speech," in Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 2014 International Conference on. IEEE, 2014, pp. 1343–1348.