International Journal of Engineering & Technology, 9 (2) (2020) 582-585

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Acoustic data classification using random forest algorithm and feed forward neural network

Ali Najdet Nasret Coran [1] *, Prof. Dr. Hayri Sever [2], Dr. Murad Ahmed Mohammed Amin [3]

[1] Cankaya university department of Electronic and Communication engineering, Ankara, turkey
[2] Cankaya university department of software engineering, Ankara, turkey
[3] Northern technical university, Mosul, Iraq
**Corresponding author E-mail: alinajdet@ntu.edu.iq*

Abstract

Speaker identification systems are designed to recognize the speaker or set of speakers according to their acoustic analysis. Many approaches are made to perform the acoustic analysis in the speech signal, the general description of those systems is time and frequency domain analysis. In this paper, acoustic information is extracted from the speech signals using MFCC and Fundamental Frequency methods combination. The results are classified using two different algorithms such as Random-forest and Feed Forward Neural Network. The FFNN classifier integration with the acoustic model resulted a recognition accuracy of 91.4 %. The CMU ARCTIC Database is referred in this work.

Keywords: *FFNN; RF; MFCC; Pitch Period; Sampling Rate; Neurons; Weight.*

1. Introduction

Speaker identification system (SIS) aims to distinguish speakers base on their acoustic information and however, speaker identification process relays on the system ability to extract the hidden pattern of the so-called speech [1]. The identification process is taking place with reference to a database where many voices are residing. The main task of the speaker identification system is to determine the level of correlation between the input signal and database. Speaker identification relies on a process called acoustic analysis and classification (matching) [2]. The acoustic analysis referred to derivation of acoustic information from the said acoustic signal by passing the signal through several stages (transfer functions) where the hidden information can be obtained. Generally, some algorithms are used while processing the voice signal and outperformed in speaker identification process. acoustic signal can be processed in either time domain or in frequency domain or in both [3]. Each domain can participate the identification procedure by adding more characteristics of the signal for accurate identification of speakers [4]. The time domain analysis can be achieved by applying cross-correlation of the signal which is very useful in determining the fundamental frequency (Pitch Period). Frequency domain analysis can be conducted after signal representation in frequency domain, frequently, discreet time Fourier transform DFT is used to convert the time domain signal in to frequency representation ,it yields the information of each frequency associated in the speech signal. Form the other hand, classification is the process to relate the particular acoustic information to their predefined speaker [5].The classification can be clearly seen when large number of speakers are allotted to the speaker identification system and hence, matching the results of the acoustic model to the particular speaker can be done seamlessly. The most deployed methods to perform classifications inspeaker Identification systems are data mining algorithms. In this paper, a text dependent speaker identification system is made. The acoustic model is constructed by determining the fundamental frequency for the input speech signals which is used as unique identity for each of the signals, the acoustic information is feed into classifiers for matching process. In this study, Random Forest Algorithm and Feed Forward Neural Network are used for classification. The experimental results shown the Feed Forward Neural Network is outperformed by achieving the higher recognition rate with less error.

2. Dataset preparation

A corpus dataset with two hundred and fifty speech signals is used in this project. The speech signals are recorded at 16000 Hz sampling rate for male English speaker. The voice signals are made with (wav) encryption and mono channel. It has been downloaded from CMU ARCTIC Databases.
The first step of preparation of the dataset is made by making the proper index of all speech signals available in the said dataset. The index is made as char array which carry the name of the respective speech signal. This array is used to loop through all the dataset contains and to simplify data splitting into test set and transit.

Copyright © Ali Najdet Nasret CORAN et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

3. Acoustic model

Sensing acoustic information which acts as real identifiers of the speech signals are obtained through the so-called Acoustic model. The dataset contents are feed to this model for determination of their pitch period and hence to calculate the Mel frequency of each utterance. In order to do so; each signal is cross correlated with its same copy for defining the top maximum peaks in their correlation resulted signal. The cross-correlation can be evaluating as per the following derivation.
Let $x[n]$ is the input speech signal, and let X[n]is the same copy of the signal. Hence m is the number of the samples in the signal. The cross correlation can be derived using the Eq. 1.[6]

$$Cor = \sum_{n=1}^{n=m} x[n] \tag{1}$$

The term Cor is shown in the Figure (1). Here, it is noteworthy to mention that the resultant signal (cross- correlation result) is dominant a double length of the original signal.

$$\mathrm{Lcor}=2.\mathrm{Lx} \tag{2}$$

The main reason of calculation the cross-correlation is to evaluate the so-called local maxima, in other word, the peak amplitude if the cross correlated signal as Figure 1 and Figure 2 depict is located at the center of the plot and represents the maximum similarity of the signals participated in correlation function Eq. (1). The local first maxima is depicted in the Figure 2

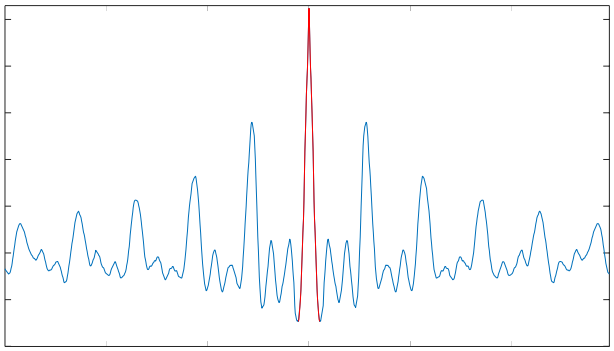


Fig. 1: Local Maxima Peak of the Cross-Correlation Function.

The process of evaluation the fundamental frequencycan be givenas:
Let Cor (fm) is the cross-correlation function at the first local maxima (firstpeak) $\mathrm{and}\,\mathrm{Cor}[sm]$is the cross- correlation function in the second local maxima (second peak). Then pitch period can be calculated as per Eq. (3) and Eq. (4), the same is depicted in Figure2.

$$|P_{laaaina} = |fm - sm| \tag{3}$$

$$|P_{leadina} = |sm - fm| \tag{4}$$

$$F_f = \frac{P^{-1}}{F_s} \tag{5}$$

Where Ff is the fundamental frequency and $\mathrm{Fs}$is the sampling frequency.

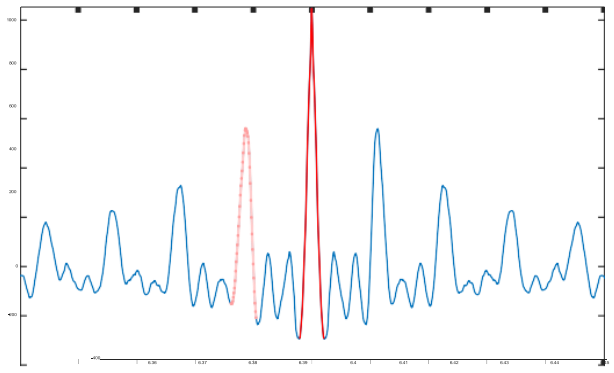


Fig. 2: A Depict of Local Maxima Is Cross-Correlation Function.

The acoustic model (paradigm) is made to evaluate the fundamental frequency for all speech signals at dataset. Eventually, a set of 250 length array is resulting which reflect the acoustic information of the whole signals in data set which is ready for classification process. Hence after, MFCC algorithm is applied to extract frequency parameters. The both results from pitch period and MFCC are merged together for accurate identification of speakers.

3.1. Random-forest algorithm

It is classification algorithm works to segregate the data into their class's base of the target information. In our case, the target is prepared to give the serial number of the speech signal as per the dataset index. The random forest is established firstly by feeding the dataset (acoustic information) along with the target into the algorithm workplace [7].
The data is feed as coma separated values which is needed to be converted into lists of float values. As soon as the data is prepared, it is spitted into number of folds to construct the trees. In our prototype, optimum number of folds was found equal to FIVE folds. As per the Random Forest rules, data is segregated into number of trees where each tree must represent the data related to same class. In other, word, tree wise data need to be of the same nature and to measure the similarity index in each tree, Gini index is calculated every time tree is constructed (training phase). Finally test data reapplied to the ultimate tree for prediction phase. The Random Forest paradigm is designed in accordance with the Figure3.

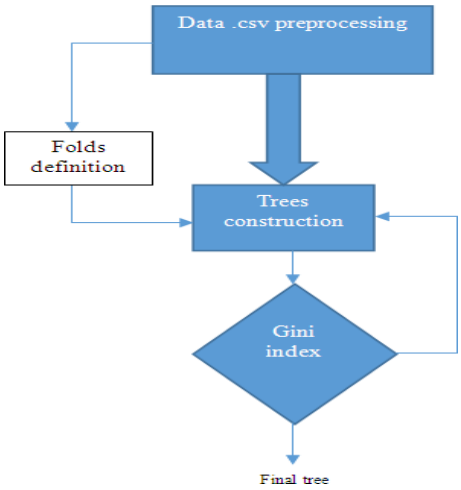


Fig. 3: Working Mechanism of Random-Forest Model.

3.2. Feed forward neural network

FFNN is popular type of neural network able to learn a complex problem of real-life applications. This kind of neural network is outperformed in learning the problems that independent of time unlikely the Current Neural Network that typically made to learn out timely problems. It is obvious, the speaker recognition process of this project and majority of similar projects are producing a time independent data so that, best type of neural network with minimal computational cost is chosen to be Feed Forward Neural network.
The model parameters are tabulated as following:

Table 1: FFNN Model parameters

Particle	Details
Number of Layers	3
Nodes (Layer Wise)	30, 10, 1
Learning method	LM
Minimum Grandniece	1e-29
Iterations	100

The model is fed with data for a very first experiment and hence all the said performance metrics are obtained. The experiment is repeated for100 times due to the random nature of weigh/bias as segment.

4. Speech classification

As the acoustic information gained from the aforementioned procedure; it is now necessary to map each fundamental frequency to its own speaker if a set of speakers willing to get identified (in testing part). Hence, the state of the art in this work is using more than one data classification algorithm to implement these requirements. Prior to get in deep with the said classifier, the data is divided in two parts as 70 percent of data is for training of the classifier and rest 30 percent of the it is for testing the performance of the trained classifier. The classification is examined using Random Forest Algorithm and Feed Forward neural Network. In each classifier, performance metrics are evaluated, so-to-say, the following terms are being calculated: Mean Square Error (MSE),Root Mean Square classifier to perform the approximation. So, those classifiers are made to learn from the acoustic information and predict the speaker identity.

5. Results and discussion

For Random-Forest and Feed Forward Neural Network classifiers, the prediction of speaker identification is performed after training process of each classifier. However, the performance metrics are evaluated for both of them and results are given as below:
For FFNN classifier, as model is suffering from random as segment of weight(*neurons)values, the results can not be fixed to a particular level, it actually gets change at every time model is restarted. For the same, 100 iterations of the experiment are performed and however the result of the accuracy can be demonstrated in Figure 4.
The comparison between the two models is given in the Table 2. Results shown that the mean FFNN is outperformed as speech classifier in speaker identification system.

Table 2: The Results of Acoustic Data classification

Classifier	MSE	RMSE	Accuracy	Time
RF	68.9	8.3006	10	6.17362
FFNN	1.23E-27	3.5071e-14	91.493	1.3248

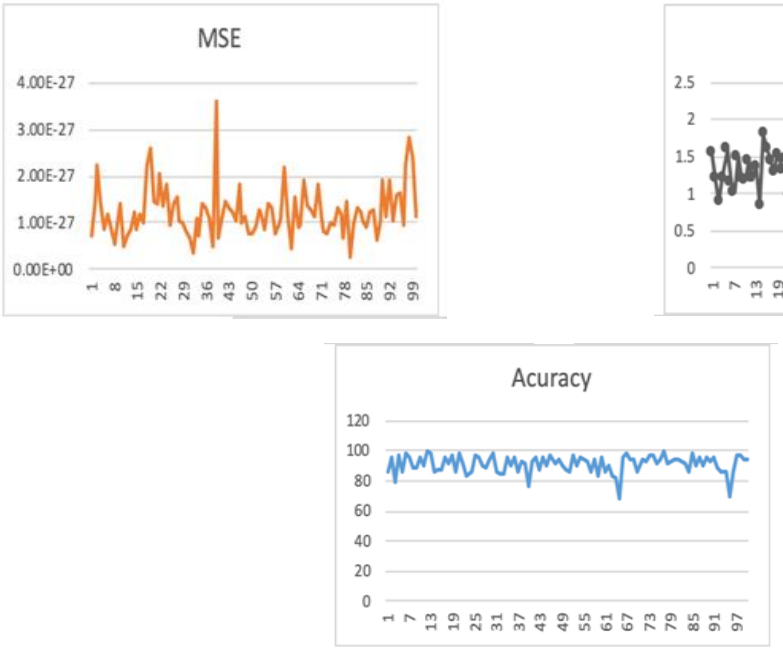


Fig. 4: The Resulted Performance of FFNN Classifier.

6. Conclusion

Speaker identification system is implemented using a combination of MFCC and Pitch Frequency models. Model is integrated with neural network and Random Forest Algorithm to predict the speaker. A text dependent speaker identification system is established ultimately. The 70 percent of the acoustic data is used to train each classifier individually and the remained 30 percent is used to test the models. The results revealed that FFNN based Speaker identification system is out performed, the same realized with 91.4 percent of recognition accuracy whereas the Random-Forest classifier has yielded only 10 percent of recognition accuracy.

References

[1] ITU Radio Regulations – Article 1, Definitions of Radio Services, Article 1.2 Administration: Any governmental department or ser- vice responsible for discharging the obligations undertaken in the Constitution of the International Telecommunication Union, in the Convention of the International Telecommunication Union and in the Administrative Regulations (CS1002).

[2] International Telecommunication Union´s Radio Regulations, Edi- tion of2012.

[3] Colin Robinson (2003). Competition and regulation in utility mar- kets. Edward Elgar Publishing. p. 175. ISBN978-1-84376-230-0.

[4] AyubiPreet, Department of Computer Science and Engineering, SGGSWU, Fatehgarh Sahib, Punjab, India (140406), AyubiPreet et al, / (IJCSIT) International Journal of Computer Science and Infor- mation Technologies, Vol. 5 (4), 2014,5508-5511.

[5] ParthaPratim Bhattacharya, RonakKhandelwal, Rishita Gera, Anjali Agarwal," Smart Radio Spectrum Management for Cognitive Ra- dio", Department of Electronics and Communication Engineering Faculty of Engineering and Technology Mody Institute of Technol- ogy and Science (Deemed University), *International Journal of Distributed and Parallel Systems* (IJDPS) Vol.2, No.4, July2011.

[6] Bablu Kumar Singh, JitendraJangir, "A Study of Recent Trends in Cognitive Radio Communications and Networks for Licence Free Connectivity", *International Journal of Engineering Research & Technology (IJERT)* ISSN:2278-0181.

[7] GoutamGhosh, Prasun Das and Subhajit Chatterjee, "Cognitive Ra- dio and Dynamic Spectrum Access a Study", *International Journal of Next-Generation Networks (IJNGN)* Vol.6, No.1, March2014. https://doi.org/10.5121/ijngn.2014.6104.

[8] Mahmood, Zuhair Shakor, Ali Najdet Nasret Coran, and Attallah Younus Aewayd. "The Impact of Relay Node Deployment In Vehicle Ad Hoc Network: Reachability Enhancement Approach." *2019 Global Conference for Advancement in Technology (GCAT)*. IEEE, 2019. https://doi.org/10.1109/GCAT47503.2019.8978445.