Digital Signal Matching Technique

Sushant S. Kalamkar
M.Sc. Student
Department of CS
Solapur University Solapur

Anuradha S.Kakade
M.Sc. Student
Department of CS
Solapur University Solapur

Niked P. Bhajibhakare
M.Sc. Student
Department of CS
Solapur University Solapur

ABSTRACT
This paper describes the Matching Model of two different voices. Here, we are working on Digital Signals & Frequency of particular Speech. We have use Mel Frequency Cepstrum Coefficients (MFCC) for matching the frequency of speech as well as used the DISTMIN for calculating the minimum distance between two different signals. This technique gives the accuracy about authorized speaker. This will shows the average of matched voice so that you can identify the speaker or voice of that speaker.

General Terms
Human computer Iterface, Matching technique, Speech Recording, Signal processing, Noise Filtering.

Keywords
Analysis, MFCC, DISTMIN, Feature extraction, Record function, Bandpass filtering.

1. INTRODUCTION
This paper shows the process of matching two different voices. We have created one model which is work digital signals and different frequencies of recorded speech. We have given one sample frequency which will detect the matched signals and converted into Frames. If coming frequency is different than sample frequency then frame will not created.

Likewise frames are created for two different voices. After that Matching purpose gives us only matched frames out of total frames of two voices in the form of Result, which will help us to identify the authorized speaker or identify the speech. Speaker identification mechanism is useful to verify identity and control of speaker.

In this paper a concept of matching two authentication modalities, the stored voice of speaker with different frequency and the speaker's runtime voice is presented that allows for flexible identification and verification with a high degree of security: a concept called Conversational Speech Biometrics [1,2]. Speech signal consist of two important types of information: first type is data content of speech and second type is the identity of that speaker. The speaker recognizes technique is used to extract identity of person [3].

Generally we are using PIN and password based systems. Now Voice passwords have been proposed but somewhere it is existing, it has work with utterance veri cation for access control and password com- pliance [7,8,9]. Biometrics and in particular speaker recognition rely for what you are. The new approach of speech biometrics or conversational biometrics [2 persons text-independent speaker recognition to acoustically identify or verify answers from the user in dialog with the system.

1.1 Preprocessing
After record new voice we have to perform filtering on it. We have use bandpass filter for remove the unwanted noise from file. After that as algorithm step it will extract features of sound file. We have show two important features first is Time and second is Average of that time. After that when you matching technique will be executed. In following fig 1 we have shown the preprocessing and algorithm of model which we have been implemented.

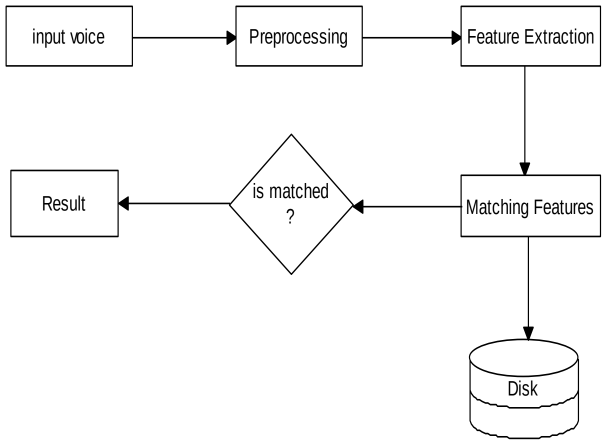

Fig 1: Preprocessing and algorithmic steps of model

1.2 Sequence of Algorithm
We have created one model in MATLAB for showing similarity within two different voices, which will show in fig 2. This gives us the path of use in system. Basically, our technique is helpful for user to record his/her voice for first stage of matching that is analysis purpose. Here users have to choose record with sampling frequency. After completing recording process, two graphs will be displayed with histograms and frequency bands. First graph will show amplitude of voice signals and second graph will show spectrogram. After that you can apply Matching techniques with accessing two different .wave saved files and check out the matched frames.

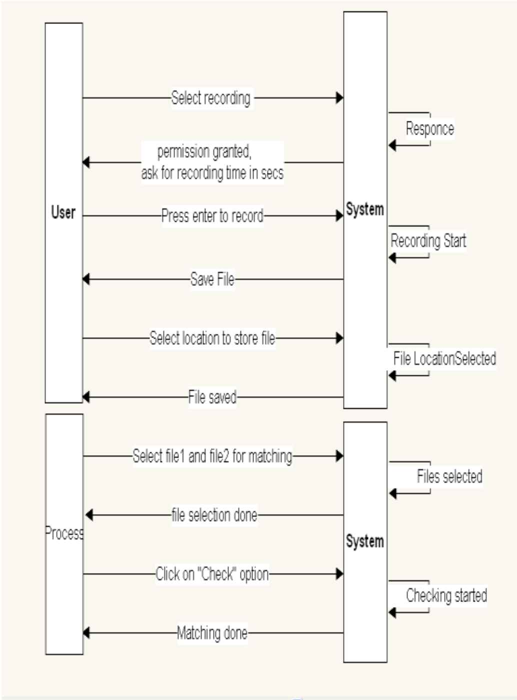

Fig 2: Sequence of Voice Matching System

2. METHODOLOGY

This paper is based on identify the speaker and match the runtime voice of that speaker with existing one. For this mechanism we have use following stages in our model for implementation.

2.1 Speech recording

This is first stage for getting the input or storing the Speech for identification afterword in matching process. We are use the technique Mel Frequency Cepstral Coefficient (MFCC) in recording function. This technique is used to create the fingerprint of the voice files. The MFCC technique is based to detect critical frequencies of bandwidth with filters spaced linearly and logarithmically at low and high frequencies. The MFCC technique is also used in detecting of various characteristics from given speech. [4], [5], [6]. Followings are some issues are overcome in recording a new speech;

Environmental Noise	**Depend on working condition**
Transducer	**Telephone, Microphone**
Amplitude of channel band	**Distortion (echo)**
Speaker of speech	**Depend or independent on sex, age or physical state**
Style of speech Voice tune	**Shouted, quite or normal**

2.2 Noise Filtering

This is next step of algorithm. In this stage we have removed unwanted noise from recorded sounds. It is very important approach because when we record some sound then it will generate some extra unwanted noise which is not required for matching. If we match sounds without apply filtering then it will give you low accuracy. But if you filter speech and match then it will give you something better accuracy.

For removing unwanted noise we have use the 'Bandpass Filter' technique which will detect all noise and remove it.

2.2.1 Bandpass Filters

Bandreject filters remove a band of frequencies about the origin of the fourier transform. An ideal bandreject filter is given by the expression,

$$H(u,v)=\begin{cases} 1 & if\ D(u,v) < D_0 - \frac{W}{2} \\ 0 & if\ D_0 - \frac{W}{2} \le D(u,v) \le D_0 + \frac{W}{2} \\ 1 & if\ D(u,v) > D_0 + \frac{W}{2} \end{cases}$$

A bandpass filler performs the opposite operation of a bandreject filler.

$$H_{bp}(u,\ v) = 1 - H_{br}(u,\ v)$$

Where Hbp(u, v) function from this expression is obtained from bandreject filters. It is transfer function.

2.3 Feature Extraction

This is the main goal that is feature extraction step of given sound is to detect parsimonious sequence of feature vectors which is used to provide better representation of input signals of sound. Feature extraction will work within three stages. The first stage analysis the speech or detect the front end. It perform spectro temporal analysis of coming signals and it generates some raw features which describing the power spectrum of intervals in speech. Second stage will compile given vectors of features into static and dynamic features. In last third stage transform vectors of static and dynamic features into better compact and robust vectors. Finally these features are supplied to recognizer [10].

In this paper we have extract two important features that is one is Time and second is Average of that Time. Those features for different two files will be stored separately. These features are so useful for matching because matching will be done with checking similarity in that features.

These features converted into frames when those are matched with given sampling frequency. One frame is created when features are matched. Likewise Frames are incremented.

We are providing one sample of frequency which is compared with new recorded file. If there are different frequency bands then frame cannot be created, but if frequencies matched to new sounds frequency then frame will be created.

2.4 Matching Technique

In this paper we have use DISTMIN and MFCC. DISTMIN is used to calculate minimum distance between coming two frequencies. MFCC is used to compare coming frequency with given sampling frequency. We have provided one sample of frequency for comparison. If coming frequency is mismatched with sampling then Frame will not created. Oppositely if frequency will match then one frame will created. Likewise frames are incremented. For matching frames we have use two features of sounds Time and Average. Result will show you matched frames out of total frames.

We have written following function code in the MATLAB for matching these frames;

```
% --- Executes on button press in pushbutton5.
function pushbutton5_Callback(hObject, eventdata, handles)
% hObject    handle to pushbutton5 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
tic;
msgbox('Checking.........','Message Box')
fs=44100;
count1=0;
count2=0;
countmatch=0;

file_name1=get(handles.txtfilename1,'String')
sound_read1=wavread(file_name1,fs*3)
fr_duration1=0.1
frame_len1=fr_duration1*fs
n1=length(sound_read1)
num_frames1=floor(n1/frame_len1)
for k=1:num_frames1
    frame=sound_read1((k-1)*frame_len1+1:frame_len1*k)
    max_val=max(frame)
    if (max_val>0.3)
        count1=count1+1
        new_sound1((count1-1)*frame_len1+1:frame_len1*count1)=frame

    end
end
```

Fig 3: MATLAB code for matching sound files

There are two techniques are available for matching approach, those are working with following ways (Svendsen et al., 1989) [11] ;

2.4.1 Whole-word matching

System compares incoming digital audio signal against a prerecorded template of the word. This technique requires less processing than sub-word matching, but it requires prerecord every word from user. As well as Whole-word templates require large amount of storage (between 50 and 512 bytes per word) because sometimes several hundred thousand words are there.

2.4.2 Sub-word matching

The system taking only sub-words and then performs further matching technique on those. This technique requires more processing than whole-word matching, but it requires much less storage (between 5 and 20 bytes per word). In addition, the pronunciation of the word can be taken from English text without requiring the user have to speak the word beforehand. (Svendsen et al., 1989) [11], (Rabiner et al., 1981) [12], and (Wilpon et al., 1988) [13] discuss that research in the area of automatic speech recognition had been pursued for the last three decades.

Only whole-word based speech recognition systems have found in the use of practical and have become better successes. Though whole word models have become a success, the researchers are agree with above techniques but they have two major problems, which is co-articulation problems and requiring a lot of training to build a good recognizer.

3. RESULT

We have been checked the experimental results with our model. We have taken two different speeches in system and apply our matching technique on these files then we have got only matched frames from files which are compared with their features. We also got as a result, how many frames are matched from total out of frames.

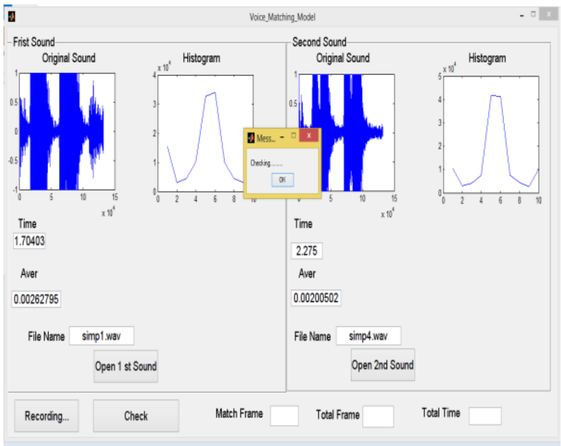

Fig 4: Checking Matching Result

We have shown Histograms and Frequency Bands separately for two different speeches. In figure 4 there is matching process have been executed. It will check matched frames from out of frames. Total frames of sounds are related with their time and average features. These are incremented when frequency of sound will match to given sample frequency.

In figure 5, actually result has displayed. Where you can see matched frames and total frames. Required time for matching process will also display. We have also removed unwanted noise from speech with apply bandpass filter. Again we show the noise removed frequency bands.

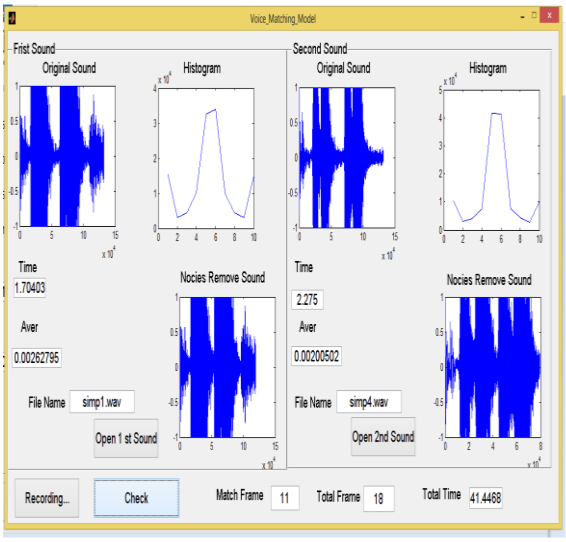

Fig 5: Result of Matching technique

We got accuracy as follow;

Speech with Noise	**40%**
Filtered Speech	**90%**
Speech with different Frequency bands	**60%**
Speech with similar Frequency bands	**80%**

4. CONCLUSION

In this paper we have provide better security as speaker identification phase, we used MFCC and Distance Minimum techniques (DISTMIN) techniques. These two techniques provided more efficient speaker identification approach. The coding of all the techniques mentioned above has been done using MATLAB. We found that the combination of MFCC and DISTMIN that is Distance Minimum algorithm gives the best performance and also accurate results.

With this matching technique you can identify speaker as well as identify any older speech with runtime speech. Which help you to check differences between older and new speeches.

5. ACKNOWLEDGEMENTS

We acknowledge Solapur University, Solapur and Department of Computer Science for the encouragement and permission to publish this paper. We would like to thank to guide Dr. Hegadi R.S. (Head of Computer Application Department) for his best guidance and support. Our special thanks to Dr. Mente R.S. (Head of Computer Science Department) for his valuable suggestions from time to time.

6. REFERENCES

[1] Maes S.H. and Beigi H.S.: Open Sesame! Speech password or key to secure your door. In Proc. ACCV, (1988), invited paper.

[2] Maes S.H.: Conversational biometrics In Proc. Of the European Conference on Speech Communication and Technology (EUROPSPEECH), Budapest Hungry, (1999).

[3] Ganesh Tiwari, “Text Prompted Remote Speaker Authentication : Joint Speech and Speaker Recognision/Verification System”.

[4] Ahsanul kabir, Sheikh Mohammad Masudul Ahsan,”Vector Quantization in Text Dependent Automatic Speaker Recognition using Mel-Frequency Cepstrum Coefficient”, 6th WSEAS Computer Science & Engineering: An international Journal (CSEIJ), Vol. 3, No. 4, August 2013 27 International Conference on circuits, systems, electronics, control & signal processing, Cairo,Egypt, dec 29-1,2007,page 352-355

[5] Lindasalwa Muda, Mumtaj Begam and Elamvazuthi, “Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and DTW Technique “,Journal of Computing, Volume 2, Issue 3, March 2010

[6] Mahdi Shaneh and Azizollah Taheri. “Voice Command Recognision System based on MFCC and VQ

[7] Algorithms”, World Academy of Science, Engineering and Technology Journal, 2009 Rosenberg A.E. and Parthasarathy S. : Speaker identification with user-selected password phrases. In Proc. Eurospeech, volume 3, pages 1371{1374,(1997).

[8] Li Q., Juange B.-H., Zhou Q., and Lee C.-H: Verbal information verification. In Proc. Eurospeech, volume 2, pages 839{842, (1997). 16. Maes S.H. and Beigi H.S.: Open Seame! Speech pas.

[9] Kimball O., Schmidt M., Gish H., and Waterman J.: Speaker Verification with limited enrollment data. In Proc. Eurospeech, volume 2, pages 967{970,(1997). 14. Lee C.-H., Soong F.K., and Paliwal K. K., editors: Automatic speech and speak.

[10] M.A.Anusuya Department of Computer Science and Engineering Sri Jaya chamarajendra College of Engineering Mysore, India, S. K. Katti Department of Computer Science and Engineering Sri Jayachamarajendra College of Engineering Mysore, India : Speech Recognition by Machine: A Review, (IJCSIS) International Journal of Computer Science and Information Security, Vol. 6, No. 3, 2009

[11] Svendsen T., Paliwal K. K., Harborg E., Husy P. O. (1989). Proc. ICASSP’89, Glasgow

[12] Rabiner, L., R., and Wilpon, J. G., (1979). Considerations in applying clustering techniques to speaker-independent word recognision. Journal of Acoustic Society of America66(3):663-673.

[13] Wilpon J. G., D. M. DeMarco, R. P. Mikkilineni (1998) “Isolated word recognision over the DD telephone network –Result of two extensive field studies”, Proc. ICASSP,pp.55-58