# Forensic Discrimination of Voices using Multi-Speech Techniques

Bhanudas K. Dethe [*1], Ajit V. Waghmare [2], Vitthal G. Mulik[3], B. P More [4], B. B. Daundkar [5]

Directorate Of Forensic Science Laboratories,kalian snatacruz (E) Mumbai, Home Dept.
Govt. of Maharashtra

1.bhanudasdethe@gmail.com, 2. .ajitwagh300@gmail.com,3.vitthalmulik96@gmail.com, 4.bhaumore1@gmail.com

**ABSTRACT:** Physiological and behavioral characteristics of voice help us to discriminate one voice from another. Uniqueness in the resonant frequencies proves that voice is unique. It has always been challenging to distinguish voices of same loudness and same pitch. In this paper, voice comparison is made on the basis of Spectrographic analysis using Multi-Speech. This technique uses Linear Predictive Coding and accordingly the formant frequencies are compared. The comparison is made purely on the basis of matching of formant frequencies of vowels.

Keywords: Forensic Science, Speech analysis, Voice analysis, Sound, Resonant frequency, Formant**.**

## INTRODUCTION

Forensic Science Laboratory, Mumbai receives number of cases for voice analysis. **We get audio recordings in two different sets, one which is recorded at the time of crime scene and the other which is taken by police from suspects. Generally recording were provided in cases such as** bribery, obscene telephone calls, kidnapping for ransom, , bomb threat calls, call girl rackets terrorist to claim credit for a terrorist action, false telephonic message etc. It becomes difficult to distinguish speakers of same loudness and pitch.

The auditory analysis (Critical Listening) and the spectrographic analysis of two sets of audio recordings are conducted to identify the similarity or dissimilarity between the voices.

In this paper, four speakers are considered to demonstrate voice recognition and to distinguish speakers of same loudness. The study begins with following steps.

## RELATED WORK

To provide the solution for the above problem there are different methodologies suggested by various people we are going to discuss some of them below.

M. Yuan, T. Lee, P. C. Ching, and Y. Zhu[1] Propose a paper Speech recognition on DSP: Issues on computational efficiency and performance analysis. In this paper provides a thorough description of the implementation of automatic speech recognition (ASR) algorithms on a fixed-point digital signal processor (DSP). It is intended to serve as a useful self-contained reference for DSP engineers to follow when developing similar applications. It is based a detailed analysis of hidden Markov model (HMM) based ASR algorithms. The computationally critical steps are clearly identified, and for each of them, different ways of optimization for real-time computation are suggested and evaluated. The trade-off among computational efficiency, memory requirements and recognition performance is illustrated quantitatively via three example systems, one for the recognition of isolated Chinese words and the other two for the recognition of English and Chinese digit strings, respectively. The paper also discusses about other techniques that can be implemented to further improve the recognition performance in real-world applications.

B.Burchard., R. Roemer, and O. Fox[2] publish a Paper on A single chip phoneme based HMM speech recognition system for consumer applications which describes a phoneme based HMM speech recognition system for use in localized and mobile consumer appliances. Unlike other systems this combined hardware/software approach does not need external components like memories etc. It is a real time application system.

Chadawan I., Siwat S. and Thaweesak Y.[4] Proposes a system Speech Recognition using MFCC which describes an approach of speech recognition by using the Mel-Scale Frequency Cepstral Coefficients (MFCC) extracted from speech signal of spoken words. Principal Component Analysis is employed as the supplement in feature dimensional reduction state, prior to training and testing speech samples via Maximum Likelihood Classifier (ML) and Support Vector Machine (SVM). Based on experimental database of total 40 times of speaking words collected under acoustically controlled room, the sixteen-ordered MFCC extracts have shown the improvement in recognition rates significantly when training the SVM with more MFCC samples by randomly selected from database, compared with the ML.

Partly motivated by the above work U. C. Pazhayaveetil [5] proposes Hardware implementation of a low power speech recognition system. In this paper presents the design, simulation and measurement results of a Hidden Markov Model (HMM) based isolated word recognizer IC with double mixtures. Table look-up technique is employed in this design. The chip operates at 20 MHz at 3.3 V. The recognition time is 0.5 s for a 50-word speech library. The speech IC has been verified with 467 test speech data and the recognition accuracy is 93.8%. A reference software recognizer using the same algorithm and speech library has a recognition accuracy of 94.2%. The new speech IC that uses a table look up to reduce the complexity of the circuit has approximately the same recognition accuracy as an ideal software recognizer.

P. Li and H. Tang[6] Proposes a  Design a co-processor for output probability Calculation in speech recognition  CHMM (Continuous Hidden Markov Model) based speech recognition algorithm, Output Probability Calculation (OPC) is the most computation-intensive part. To reduce power consumption and design cost, this paper presents a custom-designed co-processor to implement OPC. The standard SRAM interface of the co-processor allows it to be controlled by various micro-controllers. The co-processor has been implemented in standard-cell based approach and manufactured in 0.18µm UMC technology. Tested with a 358-state 3-mixture 27-feature 800-word HMM, the co-processor operates at 10MHz to meet real-time requirement. The power consumption of this co-processor is 1.6mW, and the die size is 1.18mm2.

## MATERIALS AND METHODS

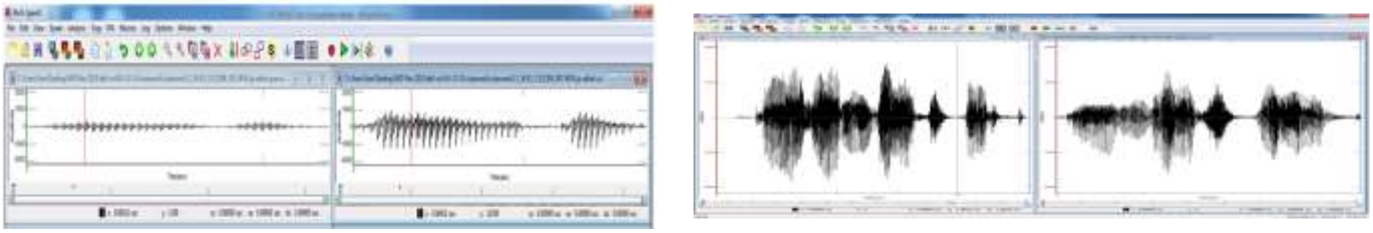
The following procedures are used for voice analysis.

**A)** Critical Listening: This is the first step performed to assess and describe the general impression of voices to be compared:  loud, dull, deep, distinct, bright, hoarse, monotonous, strong, constrained, casual, snuffling, uneducated, etc. The observation is made based on above characteristics and is noted in the observation sheet.

**B**) Segregation of Voice Samples: Audio recording which was taken from police contained more than two speakers namely complainant, suspects, witnesses. It was essential to separate out the audio of different speakers. The segregated recording of four speakers is taken for this study and is shown in fig 1.

**C)** Noise reduction: The segregated samples are then filtered with band pass filtering techniques having pass band of 400Hz to 4000Hz. This step is very important as background noise is minimized to great extent.

**D)** Spectrographic test using Multi-Speech: Spectrographic test is conducted for detailed examination of resonances. Thus the third step of voice analysis performed is spectrographic test. The spectrographic test is divided in three parts: a) Amplitude vs. Time analysis (waveform analysis) b) Frequency vs. Time analysis (formant analysis) c) Energy vs. Frequency (LPC analysis).



Amplitude vs. time of Speaker  I Amplitude vs. time of Speaker II    Amplitude vs. time of Speaker III Amplitude vs. time of Speaker IV

Fig 1 Segregated speech  of four Speakers (waveform pattern)

**Using multi speech, spectrographs** of Speaker I, Speaker II, Speaker III and Speaker IV are produced **(see Fig 2 and Fig 3)**. The formant frequencies (see table 2 and table 3) are then noted and voices are compared. The voice comparison is made based on the matching of formant frequencies.

### A SPECTROGRAPHIC RESULT (SEE FIG 2) OF SPEAKER I AND SPEAKER II SHOWS THE SIMILARITY BETWEEN THEIR VOICES AS THE FORMANT FREQUENCIES ARE MATCHING (SEE TABLE 2).

Word: /JaUdhar/

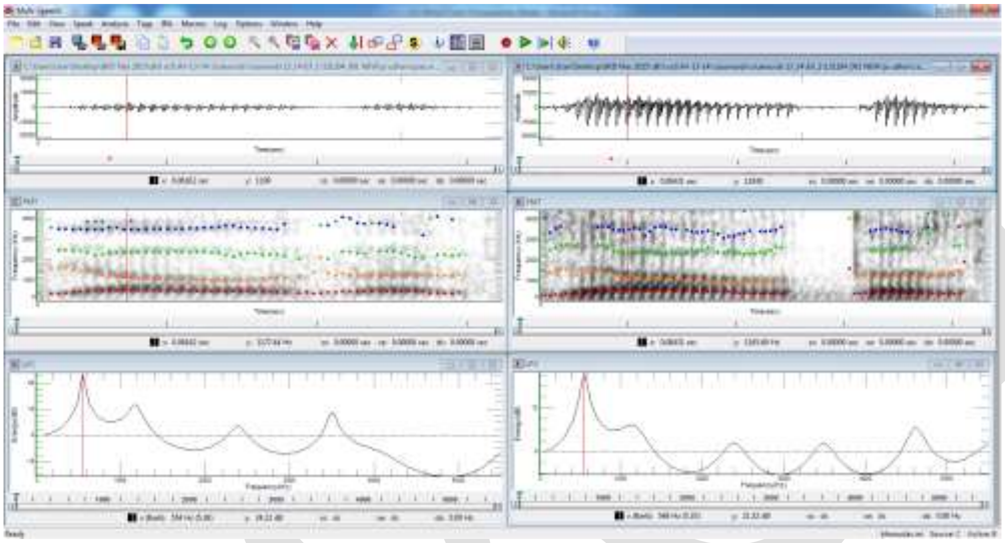
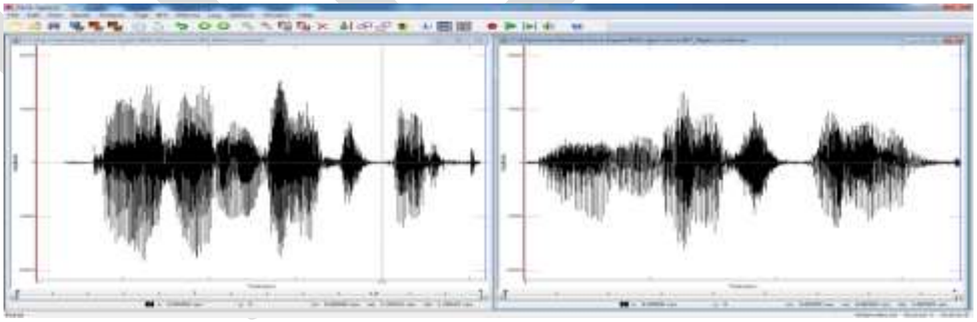


Fig 2: Spectrographic results showing similarities in speech data of speaker I (left side) and speaker 2 (right side) based on formants matching, f1, f2, f3, f4.(Analyzed in Tape Authentication and Speaker Identification Division, Forensic Science Laboratory, Mumbai, Maharashtra State)

Also the following spectrographic result (see fig 3) of speaker 3 and speaker 4 shows the dissimilarity between their voices as the formant frequencies are not matching.(see table 3).

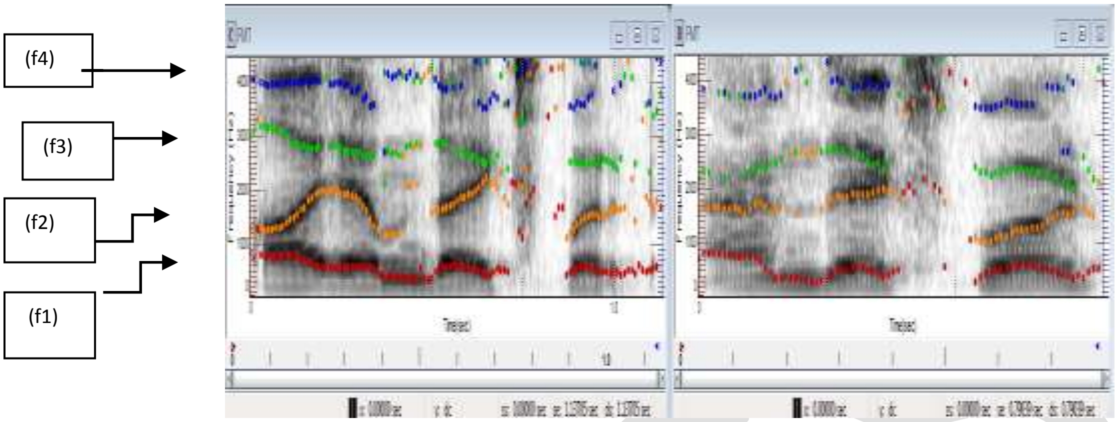
Word: *I am the expert*



Amplitude vs. time of speaker3Amplitude vs. time of speaker 4

*Fig. b*

Figure 3: Spectrographic results showing dissimilarities between speaker 3(left side) and speaker 4 (right side)**a)** waveform pattern **b)** Formant pattern (f1), (f2), (f3), (f4).(Analysed in Tape Authentication & Speaker Identification Division, Forensic Science Laboratoy, Mumbai, Maharashtra state.)



The analysis is done with following configuration.

TABLE 1

FORMANT ANALYSIS CONFIGURATION

| Analysis | Framing | Display | Colour |
|---|---|---|---|
| Filter order: 12 | Frame Length: 25msec | Frequency display: Linear | F1: Red |
| Pre-emphasis: 0.900<br>Range( 0-1.5) | Frame Advance: 5msec | Formants with bandwidth <500Hz | F2: Orange |
| Analysis method: Autocorrelation | - | Only voiced frames | F3: Green |
| Window Weighting: Blackman | - | % Nyquist: 0 to 80 % | F4: Blue |

## OBSERVATIONS

In the auditory analysis, it is observed that the sounds produced by speaker 1 and speaker 2 have low loudness whereas those of speaker 3 and speaker 4 have high loudness. Also the speakers 1 and 2 are quite uneducated whereas speaker 3 and 4 are educated.

The voice comparison can also be made based on waveform pattern (see fig 1) and formant pattern (see fig 2 and fig 3). The formant frequency pattern f2 of speaker 3 is curved (see fig3) and that of speaker 4 is flattened. It shows that two speakers are different. Also formant frequency pattern f1 and f2 of speaker 1 and speaker 2 are similar. It shows that two speakers are same. Hence speakers of same loudness can be distinguished based on formant pattern.

## RESULTS & DISCUSSION

The formant frequencies have been observed and also the bandwidth is noted. The formant frequencies of speaker 1 and speaker 2 are shown below.

TABLE 2 : LPC RESULTS

| Speaker 1 | | Speaker 2 | |
|---|---|---|---|
| Formants (Hz) | Bandwidth (Hz) | Formants (Hz) | Bandwidth (Hz) |
| **554.82** | 28.50 | **548.30** | 68.14 |
| **1177.10** | 125.49 | **1165.58** | 307.88 |
| 2380.77 | 178.72 | 2380.85 | 276.04 |
| 3494.28 | 89.64 | 3472.17 | 267.71 |
| 4033.24 | 592.82 | 4594.72 | 193.13 |

The formant frequencies of speaker 1 and speaker 2 are matching with following tolerance band. (f1)-10Hz, (f2)-20Hz, (f3)-40Hz, (f4)-80Hz. Hence two speakers are similar. The formant frequencies of speaker 3 and speaker 4 are shown below.

TABLE 3 : LPC RESULTS

| Speaker 3 | | Speaker 4 | |
|---|---|---|---|
| Formants (Hz) | Bandwidth (Hz) | Formants (Hz) | Bandwidth (Hz) |
| 664.03 | 71.03 | 820.83 | 123.06 |
| 1493.76 | 3662.21 | 1666.96 | 152.40 |
| 1854.88 | 66.98 | 2231.01 | 649.52 |
| 2773.79 | 91.75 | 3816.25 | 407.35 |
| 4009.99 | 71.61 | 4829.88 | 232.77 |

The formant frequencies of speaker 3 and speaker 4 are not matching. Hence two speakers are different(Not Similar).

Thus, two musical sounds may differ from one another in three ways. They may differ in loudness, pitch and quality or timbre.

A) Loudness: Loudness is the characteristic by virtue of which we distinguish two sounds of same frequency. It depends upon the intensity of vibration. The energy of particle performing vibrations is proportional to the square of amplitude and frequency (E œ$A^2$ $n^2$). Hence for given frequency, intensity depends upon the square of the amplitude of vibration. Thus it shows that greater is the extent of vibration i.e. amplitude, the louder is the sound it gives out. As the sound propagates through medium, there is a decrease in the amplitude i.e. sound becomes less loud. It also depends upon the following factors: i) the density of air. ii) The velocity and direction of wind. iii) The sensitivity of the ear, as the sensitiveness varies with pitch. iv)Loudness decreases with distance. v) Loudness can be increased by resonance in the presence of sounding resonators or boxes.

*B)* Pitch: Pitch of note is the characteristic that differentiates the notes. Pitch of note is the frequency of vibrations of the source and is equal to the number of vibrations performed by the source per second.
The greater is the frequency, the higher is the pitch and when the pitch is higher, the sound is said to be shrill. Notes of lower pitch are flat. Male voice is flat while that of female is shrill on account of higher pitch.

*C)* Quality or timbre: It is the characteristic that distinguishes two sounds of same pitch and loudness.

## CONCLUSIONS

Thus we can compare voices of two unknown speakers using Multi speech techniques. Two musical sounds may differ from one another in three ways namely loudness, pitch and quality or timbre. Two speakers can speak with same loudness and pitch but the third characteristic of sound i.e. The formant frequency plays very important role in distinguishing human voice.

## ACKNOWLEDGEMENT

The authors are thankful to the Forensic Science Laboratory, Mumbai Maharashtra State for providing research environment and tools used in the voice analysis.

## REFERENCES:

[1] M. Yuan, T. Lee, P. C. Ching, and Y. Zhu, “Speech recognition on DSP: Issues on computational efficiency and performance analysis,” in *Proc. IEEE ICCCAS*, 2005.
[2] B.Burchard., R. Roemer, and O. Fox, “A single chip phoneme based HMM speech recognition system for consumer applications,” *IEEE Trans. Consumer Electron.*, vol. 46, no. 3, Aug. 2000.
[3] U. C. Pazhayaveetil, “Hardware implementation of a low power speech recognition system,” Ph.D. Dissertation, Dept. Elect. Eng., North Car-olina State Univ., Raleigh, NC, 2007.
[4] Chadawan I., Siwat S. and Thaweesak Y. “Speech Recognition using MFCC”.International Conference On Computer Graphics, Simulation and Modeling (ICGSM'2012)July 28-29, 2012 Pattaya (Thailand).
[5] W. Han, K. Hon, Ch. Chan, T. Lee, Ch. Choy, K. Pun and P. C. Ching, “An HMM-based speech recognition IC,” in *Proc. IEEE ISCAS*, 2003.
[6] P. Li and H. Tang, “Design a co-processor for output probability Calculation in speech recognition,” in *Proc. IEEE ISCAS*, 2009.