Study of Variations in Mean Fundamental Frequency and Mean Energy of the Voice Recording of a Speaker with Respect to the Distance

Shital k. Parmar[a], Mrs. Surbhi Mathur[b*], Dr. M.S.Dahiya[c], Dr. J. M. Vyas[d]

[a.] *Student, IFS, Gujarat Forensic Sciences University Gandhinagar, Gujarat, India*
[b.*] *Assistant Professor Jr., Gujarat Forensic Sciences University, Gandhinagar Gujarat, India (Corresponding author)*
[c.] *Director, IFS, Gujarat Forensic Sciences University, Gandhinagar, Gujarat, India*
[d.] *Director General, Gujarat Forensic Sciences University, Gandhinagar, Gujarat, India*

***Abstract*:-Speaker recognition has become an important tool in crime investigation, since the use of the human voice as an instrument in commission of crime is ever-increasing due to increase in communication. We can identify a person on the basis of acoustical analysis of his/her voice but as we all know there are some limitations against expert in forensic speaker recognition and for this reason the voice evidence is usually accepted as the secondary evidence supporting other scientific facts in the court of law, but in offenses such as kidnapping, extortion, Blackmail threats, obscene calls, harassment calls, ransom calls etc. where the criminal catch aid of telephone or mobile phone for accomplishing their illicit tasks, voice automatically becomes the primary and most valuable evidence for linking the suspect and the crime. One limitation against expert in forensic speaker recognition is the variation in distance between recorder and the mouth of speaker during questioned voice and control voice recording.In this paper, we present a study of variations in mean fundamental frequency and mean energy of the voice recording of a speaker with respect to the distance. We analyzed the range of variation in energy and fundamental frequency for above situation which may further useful to forensic expert to overcome the limitations and achieve precise and accurate results. The results show a very strong influence of the distance variation between speaker and recording device upon mean energy.**

Keywords: - Forensic Science, Speaker Recognition Technology, Voice, Speech, Mean Fundamental Frequency, Energy.

I. INTRODUCTION

Speech is frequently cited as a most important human faculty and sometimes as a uniquely human faculty. Forensic speaker recognition has become an important tool in crime contravention, since the use of the human voice as an instrument in commission of crime is ever-increasing. Much of information is exchanged between two parties in telephone conversations, including between criminals[1,2,3].

1.1. Voice as a Biometrics

Speech is a coordinated effort of lungs, larynx, vocal cords, tongue, lips, mouth and facial muscles, all which are activated by brain. Voice quality depends on number of anatomical features such as dimension of oral tract, pharynx, nasal cavity, shape and size of tongue and lips, position of teeth, elasticity of the organs and density of vocal cords[2,4]. Each person possesses a unique voice quality, arising out of individual variations in vocal mechanism. The uniqueness in the voice lies in the fact that experimentally develops an individual and unique process of learning to speak. Chances that the two individuals will have same shape and size of vocal cavity, vocal tract, and will control their articulators in same manner is quite small.Speaker recognition has a history dating back some four decades and uses the acoustic features of speech that have been found to differ between individuals[5,6,7]. These acoustic patterns reflect both anatomy (e.g., size and shape of the throat and mouth) and learned behavioural patterns (e.g., voice pitch, speaking style).

1.2. Voice / Speaker Recognition

Speaker recognition is defined as any activity in which a speech sample is linked to a particular speaker on the basis of its acoustic or perceptual properties[8] (as seen in figure 1).
There are two major applications of speaker recognition:

1) *Voice - Speaker Verification / Authentication*

The use of the voice as a method of determining the identity of a speaker for access control. If the speaker claims to be of a certain identity and the voice is used to verify this claim. Speaker verification is a 1:1 match where one speaker's voice is matched to one template (also called a "voice print" or "voice model").

2*) Voice - Speaker Identification*

Identification is the task of determining an unknown speaker's identity. Speaker identification is a 1: N match where the voice is compared against N templates. Speaker identification systems can also be implemented covertly without the user's knowledge to identify talkers in a discussion, alert automated systems of speaker changes, check if a user is already enrolled in a system, etc.

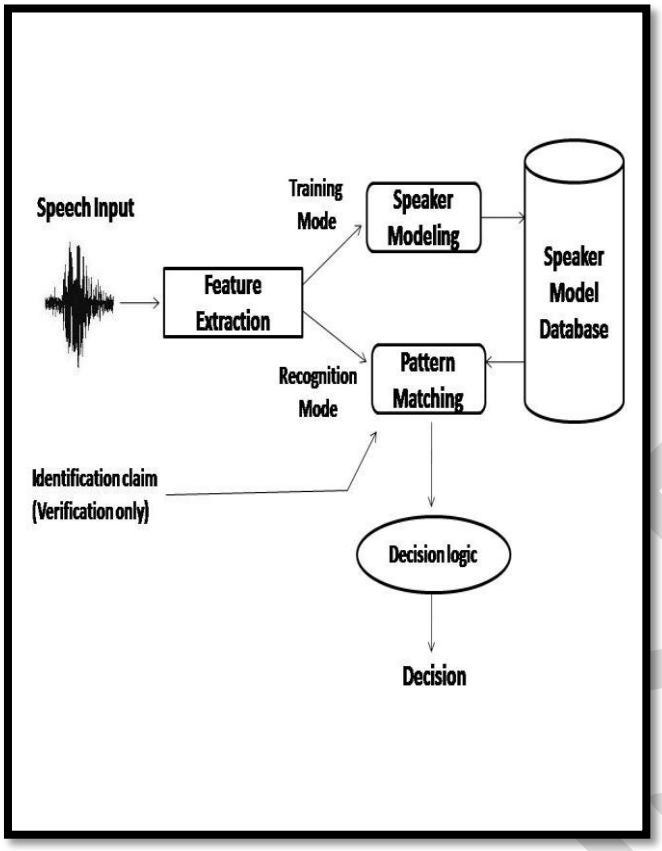


Fig. 1 Components of Speaker Recognition

To a certain degree, this is undoubtedly that voice is good biometric of a person. We can identify a person on the basis of acoustical analysis of his/her voice but as we know voice of human is behavioral biometrics hence, it causes some limitations against expert.

1.3. Limitation in Forensic Speaker Recognition

- Short duration samples are more demanding and thus should be analysed carefully.
- The dissimilar language in questioned and specimen are difficult to analyze.
- Emotion Variability in questioned and specimen samples[9].
- Misspoken or misread prompted phrases.
- Poorly recorded/noisy samples are difficult to analyze[10].
- Some changes occur in features of voice due to variation in distance between mouth of speaker and recording device[11, 12].
- Due to variation in posture of person
- Insufficient number of comparable words.
- Disguise in speech samples poses a problem in speaker recognition and/or the degree of disguise is decided by the expert.
- Extreme emotional states[13,14,15] (e.g. stress or duress)
- Change in physical state of the speaker (e.g. eating, effect of ethanol, etc).
- The attitude of the how the speech is said by the speaker[16].
- Channel mismatch or mismatch in recording conditions (e.g. using different microphones for enrolment and verification Different pronunciation speed of the test data compared with the training data.
- Speaker's health.
- Aging of the vocal tract[17]. (Vocal tract can drift away from models with age).

II. METHODOLOGY

2.1. Sample Collection

Seventeen adults (8 females and 9 males, 24 to 28 years), participated in the following experiment
,who have no history of speaking diseases. Each speaker was requested to give 7 voice samples in different recording condition. (As shown in Table No.1)

Sample No.	Recording Condition
Sample 1	Indoor, Sitting, Distance from mouth of speaker to recorder:- Average 17 Cm
Sample 2	Indoor, Sitting, Distance from mouth of speaker to recorder:- Average 40 Cm
Sample 3	Indoor, Sitting , Distance from mouth of speaker to recorder:- Average 67 Cm
Sample 4	Indoor, Standing, Distance from mouth of speaker to recorder:- Average 35 Cm
Sample 5	Indoor, Standing, Distance from mouth of speaker to recorder:- Average 75 Cm
Sample 6	Indoor, Standing, Distance from mouth of speaker to recorder:- Average 86 Cm
Sample 7	Indoor, Standing, Distance from mouth of speaker to recorder:- Average 125 Cm

Table 1: Different conditions under which the recording of the voice samples taken place

We have taken recordings at a same time in two different recording devices to cross check whether the variation in the features of speech samples, due to change in posture and distance from speaker to recorder is same in both recording device or not and to verify whether the variation is due to device or not.

Two recording device have been used for recording purpose.

1).Digital recorder

2).Gold wave software

2.2. Analysis

After collecting the voice sample from particular individual, we have analyzed the each voice sample separately and compared it. To analyze the sample we have used two different approaches complementary.

- Auditory Analysis
- Instrumental analysis

2.2.1. Auditory Analysis

For the auditory analysis, we have heard the voice of person repeatedly and analyzed the following parameters after carefully listening.

→ Quality Of Speech
→ Flow of Speech
→ Speed
→ Plosive Formation
→ Nasality
→ Intonation Pattern
→ Dynamic Loudness
→ Pauses
→ Speech Rate
→ Sample Duration
→ Number & Length of Pauses

2.2.2. Instrumental Analysis

In instrumental analysis, we have used spectrographic analysis approach with the help of software named Gold wave Software and CSL 4500(as seen in figure 2).With the help of Gold wave, we can record the voice of person as well as resample the voice sample in particular format that is Wave, 32 kbps, 44 KHz, Mono.CSL 4500(Computer Speech Lab) is a highly advanced acoustic analysis system in which we obtain a spectrographically representation and statistical value of energy, pitch and formants of voice sample.

Following parameters were taken in to notice for instrumental analysis (in CSL 4500) purpose.

- Energy
- Pitch/Fundamental frequency
- Formants Frequency

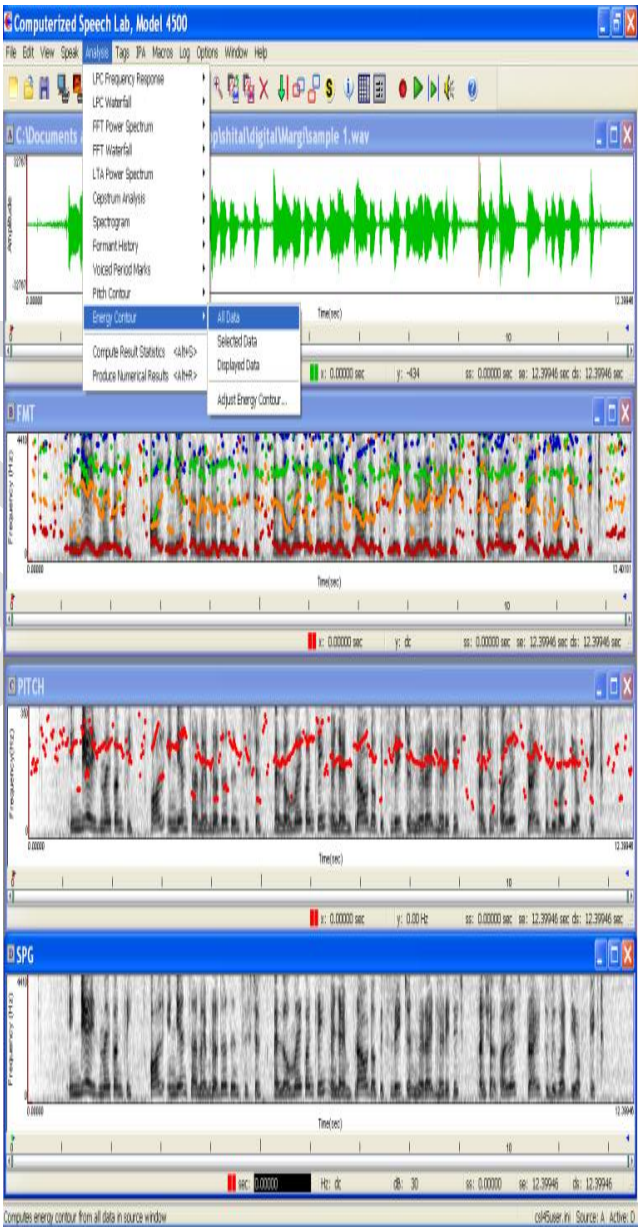


Fig.2 Spectrograpicaly representation of energy,pitch and formants in CSL 4500

-First of all, we had resample the recording sample at 16 bit, mono PCM, 11025 KHz. (With the help of gold wave software)

III. OBSERVATION

Subject No. : 1
Name: Shital Parmar
Age. 23 Yr
Sex: Female
Background: Educational-M.Sc.Forensic Science, Residential – Mangrol,Junagadh
Language used: English

Observations for auditory analysis of voice samples (For both Gold wave and Digital Recorder)

Table 2: The Speaker begins with the following words

Sample 1	India is my country……
Sample 2	India is my country……
Sample 3	India is my country……
Sample 4	India is my country……
Sample 5	India is my country……
Sample 6	India is my country……
Sample 7	India is my country……

Table 3: Recording mode

Sample No.	Digital recorder	Gold wave software
Sample 1 to 7	WMA ,32 kbps,44KHz,Mono	Wave PCM Signed 16 bit,44KHz,1411 Kbps ,Stereo

Table 4: Quality of speech samples

Sample 1	Nor mal
Sample 2	Nor mal
Sample 3	Nor mal
Sample 4	Nor mal
Sample 5	Nor mal
Sample 6	Nor mal
Sample 7	Nor mal

Table 5: Articulatory parameters of speech

Acoustic parameter Sample no.	Speech duration (In Sec)	Flow of Speech	Speech rate
Sample 1	13.600	Nor mal	23 Words/8 Sec
Sample 2	13.299	Nor mal	24 Words/8 Sec
Sample 3	12.446	Nor mal	26 Words/8 Sec
Sample 4	12.446	Nor mal	25 Words/8 Sec
Sample 5	12.074	Nor mal	26 Words/8 Sec
Sample 6	13.003	Nor mal	25 Words/8 Sec
Sample 7	10.907	Nor mal	25Words/8 Sec

Table 6: Prosodic Analysis

Acoustic parameter Sample no.	Intonation Pattern	Dynamic Loudness
Sample 1	Nor mal	Normal
Sample 2	Nor mal	Normal
Sample 3	Nor mal	Normal
Sample 4	Nor mal	Normal
Sample 5	Nor mal	Normal
Sample 6	Nor mal	Normal
Sample 7	Nor mal	Normal

Observations for instrumental analysis of voice Sample

Sample no.	Recording condition	Mean F0	Mean Energy
1	Indoor, Sitting, Near Dis. Of Mouth from Recorder :- 12Cm	224.76	70.40
2	Indoor, Sitting, Near, Dis. Of Mouth from Recorder :-30Cm	229.17	70.47
3	Indoor, Sitting, Near Dis. Of Mouth from Recorder:- 60Cm	219.67	67.32
4	Indoor, Sitting, Far Dis. Of Mouth from Recorder:- 26Cm	221.29	70.97
5	Indoor, Sitting, Far Dis. Of Mouth from Recorder :-70Cm	213.06	67.56
6	Indoor, Sitting, Far Dis. Of Mouth from Recorder :-77Cm	221.87	64.48
7	Indoor, Sitting, Far Dis. Of Mouth from Recorder :-125Cm	228.99	62.43

Table No.7: Values of mean fundamental frequency and mean energy for voice recordings using digital recorder

Sample no.	Recording condition	Mean f0	Mean energy
1	Indoor, Sitting, Near Dis. Of Mouth from Recorder :- 12Cm	224.86	56.32
2	Indoor, Sitting, Near, Dis. Of Mouth from Recorder :-30Cm	221.03	55.87
3	Indoor, Sitting, Near Dis. Of Mouth from Recorder:- 60Cm	222.74	50.25
4	Indoor, Sitting, Far Dis. Of Mouth from Recorder:- 26Cm	222.15	56.03
5	Indoor, Sitting, Far Dis. Of Mouth from Recorder :-70Cm	219.31	47.63
6	Indoor, Sitting, Far Dis. Of Mouth from Recorder :-77Cm	220.21	45.80
7	Indoor, Sitting, Far Dis. Of Mouth from Recorder :-125Cm	225.18	44.32

Table No. 8: Values of mean fundamental frequency and mean energy for voice recordings using Goldwave software

IV. RESULT AND DISCUSSION

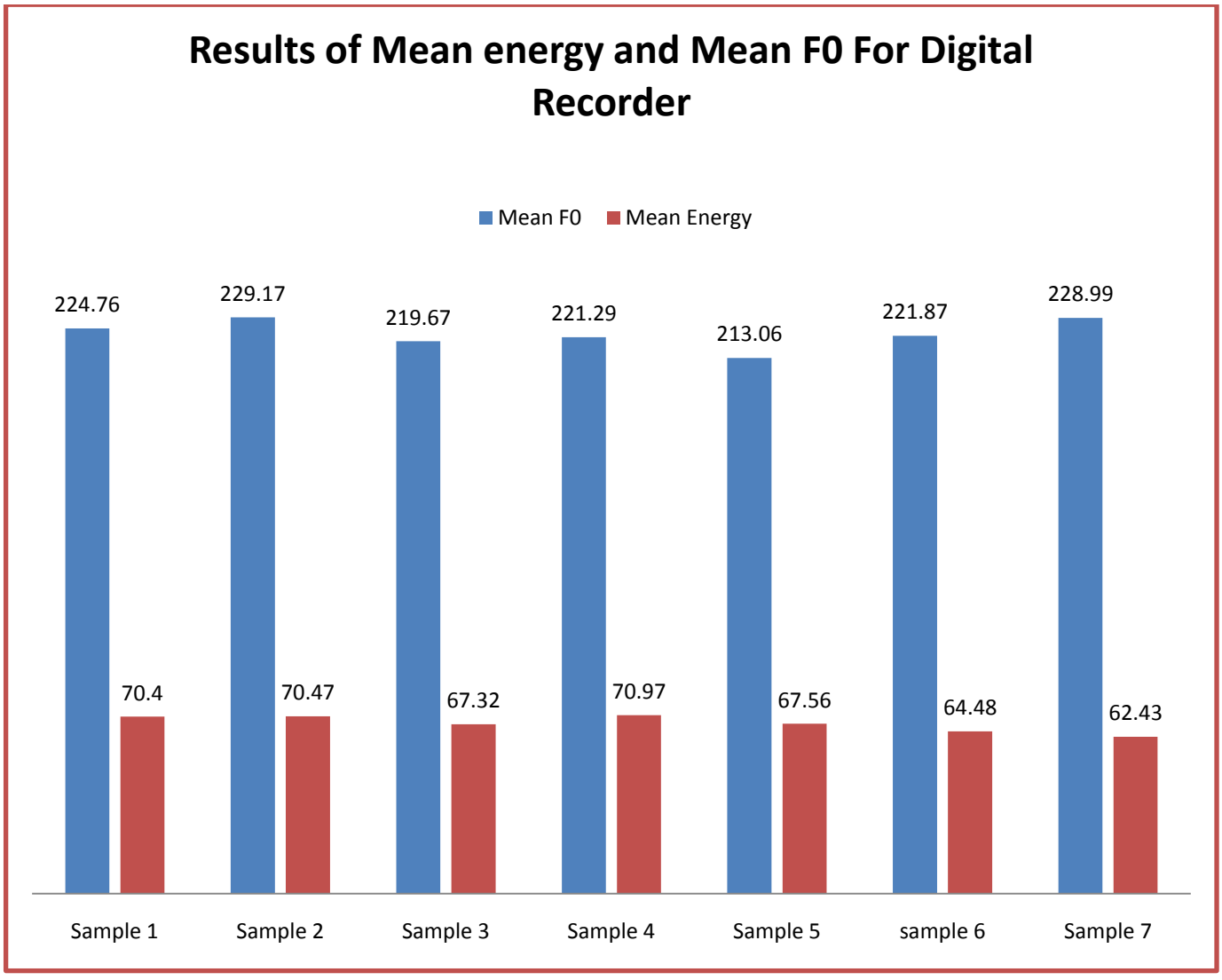
The auditory analysis and subsequently acoustic analysis by using computerized speech lab.(CSL 4500) revealed that features of voice (Mean F0 and Mean Energy)of the speaker in different recording condition vary with respect to distance between speaker and recording device.

The mean Fundamental frequency changes are not regular and do not have uniformity. The main reason for changing the fundamental frequency is a combine effect of the natural variation, speaker's emotional state, recording device, microphone, and orientation of microphone rather than change in distance between speaker and recording device.

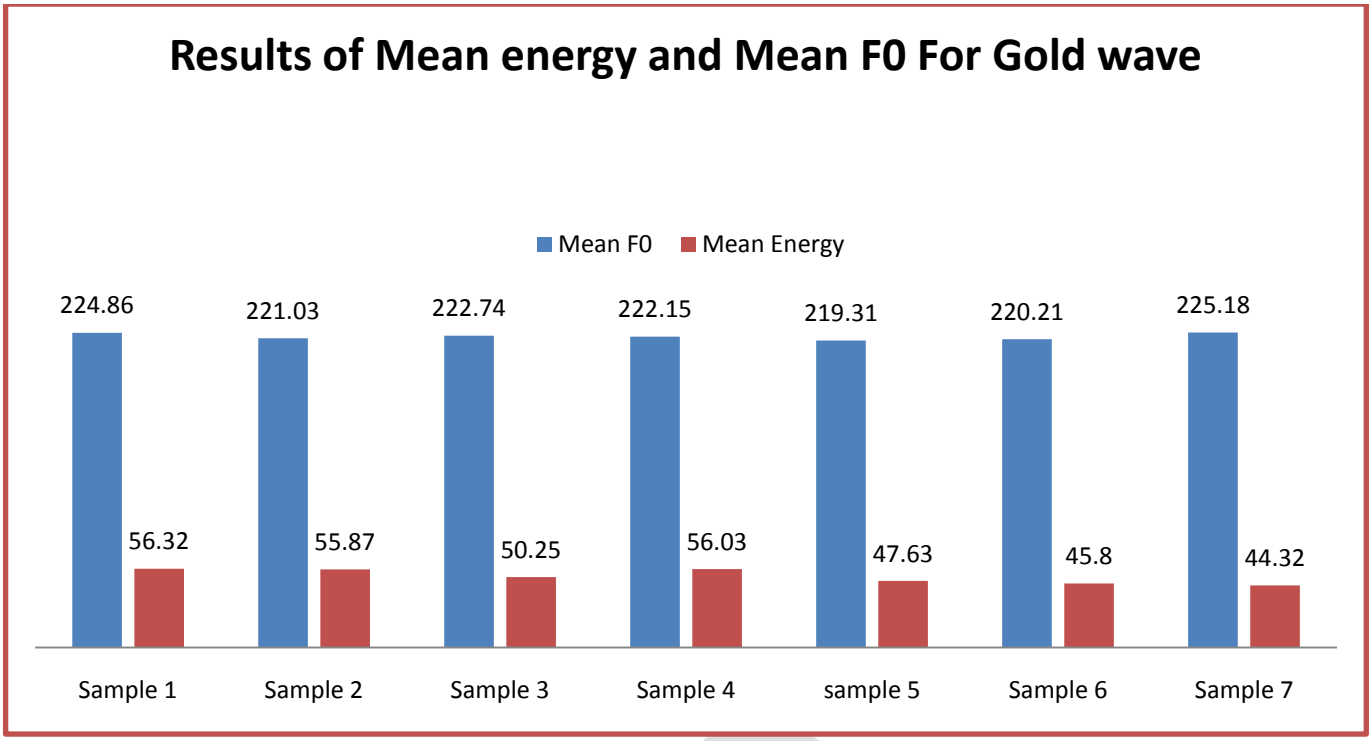
So the mean fundamental frequency shows very weak or null relationship with distance between recording device and mouth of speaker.

Energy based results gives high probability of identifying the distance between the mouth of a speaker and recording device. In the case of energy, there is a uniform similarity that the distance between speaker and recording device increase the mean energy decrease and if the distance decrease the mean energy increase during whole experiment for all 17 speaker. The graphical representations of the results of mean energy and mean F0 for both gold wave and digital recorder are given below.

As we know, the analysis of voice samples is tedious and time consuming process so we had taken limited number of samples to finish this work within given limited time. But for future work we will collect ample number of samples from different distance by increasing the distance between recorder and speaker .In our future work, we will also repeat our experiments



Graph No.1:-Results of Instrumental Analysis (Recording through Digital Recorder)



Graph No.2:- Results of Instrumental Analysis (Recording through Gold Wave Software)

by using more sophisticated recording devices and microphones.Additionally we will also collect long duration speech sample for getting more information about auditory acoustic parameters.

For future work we will also change the intensity of speaker's voice parallel to the change in distance during collection of speech samples.

There are many reasons for changing the features of voice of speaker for e.g. Variation in physical and mental condition of speaker, vocal aging, due to recording device, mental stress, due to long time speaking, body weight, according to time of day(in early morning, late night) to speak .For future work, we will also consider such conditions and additional parameters.

V. CONCLUSION

This project addresses the issue of variation in the mean F0 and mean energy of the voice recording of a speaker with respect to the distance between recorder and mouth of the speaker.

Using the collected speech database with speaker speech recordings while far and near distance from speaker to recording device, we analyzed the features of speaker's voice like meannF0 and Mean energy.

The analysis showed that there is no significant dependency between the variations in Mean F0 and distance between speaker and recorder. The mean Fundamental frequency changes are not regular and do not have uniformity. Fundamental frequency depends on the vibration of vocal folds so there are no significant changes in mean F0 due to variation in distance between speaker and recorder.

From the energy based results, we can conclude that there is significant dependency between the variation in mean energy and the distance between speaker and recording device. If the distance between speaker and recording device increase the mean energy decrease and if the distance decreases the mean energy increase during whole experiment.

REFERENCES

[1]. Eric Keller; Fundamentals of speech synthesis and speech recognition: basic concepts, state of the art, and future challenges; John Wiley and Sons Ltd. Chichester, UK; 1994
[2]. Philip Rose; Forensic Speaker Identification , Hardcover – July 1, 2002; ISBN-13: 978-0415271820
[3]. Raichel Daniel R.; The science and application of acoustics; Hardcover – May 11; ISBN-13: 978-0387989075
[4]. Harry Hollien; Forensic Voice Identification; Hardcover – October 9, 2001; ISBN-13: 978-0123526212
[5]. Ray D. Kent, Charles Read The Acoustic Analysis of Speech; Paperback – December 21, 2001; ISBN-13: 978-0769301129
[6]. Neustein, Amy, Patil and Hemant A.; Forensic Speaker Recognition: Law Enforcement and Counter-Terrorism; Hardcover – Import, 6 Oct 2011
[7]. Molly L. Erickson et al: " A Comparison of Two Methods of Formant Frequency Estimation for High-Pitched Voices"Published in Journal of voice, volume 16, issue 2, June 2002, page 147-171
[8]. Surbhi Mathur,et al: "Speaker Recognition System and Its Forensic Implications: A Review, Volume III, Issue IV, April 2014 IJLTEMAS ISSN 2278 - 2540
[9]. Marco Guzman et al: "Influence on Spectral Energy Distribution of Emotional Expression"; Journal of voice, volume 27, issue 1, January 2013, page 129.e1-129.e.

[10]. Dimitar D. Deliyski et al: "Adverse Effects of Environmental Noise on Acoustic Voice Quality Measurements"; Journal of voice, volume 19, issue 1, March 2005, page 15-28
[11]. Erkki Vikman et al; "Effects of prolonged oral reading on F_0, SPL, subglottal pressure and amplitude characteristics of glottal flow waveforms"; Journal of voice, volume 3, issue 2, June 1999, page 303-312
[12]. Theresa A. Burnett et al: "Voice F_0 responses to pitch-shifted auditory feedback: a preliminary study" Journal of voice, volume 11, issue 2, June 1997, page 202-211
[13]. Sungbok Lee, et al: "Study of Acoustic Correlates Associated with Emotional Speech"; 148th ASA Meeting, San Diego, CA
[14]. Lieberman,et al: "Fundamental frequency of phonation and perceived emotional stress"; J Acoust Soc Am. 1997 Apr; 101(4):2267-77.
[15]. Marius Vasile, et al.: "A Study of the Effect of Emotional State upon the Variation of the Fundamental Frequency of a Speaker"; Journal of Applied Computer Science & Mathematics, no.7 – Special Issue.
[16]. Sirisha Duvvuru et al:"The Effect of Change in Spectral Slope and Formant Frequencies on the Perception of Loudness"; Journal of voice, volume 27, issue 6, November 2013, page 691-697
[17]. S.Mwangi,et al.: "Effect of Vocal Aging on Fundamental Freque ncy and Formants"; NAG/DAGA 2009-Rotterdam

LIST OF FIGURES

LIST OF TABLES

LIST OF GRAPHS