ELK

Asia Pacific Journals

www.elkjournals.com

....

HUMAN VOICE ESTIMATION BASED ON DIFFERENT AGE GROUP

Shruti Agarwal	**Ashish Agrawal**	**Prabhakar Gupta**
M.Tech Dept. of CSE	Asst. Professor, Dept. of CSE	Dean Academics
SRMSCET	SRMSCET	SRMSCET,
Bareilly, India	Bareilly, India	Bareilly, India
shrutiagarwal235@gmail.com	agarwal.ashish01@gmail.com	prabhakar.gupta@srmscet.edu

Abstract

Speech authentication techniques for different ages are difficult to recognize. In this paper we discuss about the human speech recognition through different parameters. Human speech is very essential part for communicating with others. There are many parameters for recognizing the human speech according to their age group such as pitch, correlation, etc but we mainly focus on three parameters i.e. Maximum frequency of the speech, standard deviation, and the covariance of the speech signal. Speech recognition is very powerful technique that allows the people to communicate with each other through different technical gadgets.

Keywords—Fuzzy Logic(FL), MF(Maximum Frequency), CV(Coefficient of variation), Fuzzy If-Then Rules, Standard Deviation(SD), Oscilloscope

I. INTRODUCTION

Human speech is very important because with the help of speech we can interact with each other. Speech is the process through which the persons can talk with each other interact with each other. Speech is important to interact with the people but on the other side there are many people they don't have their voice, but they also interact with other human beings. There are many persons who are unable to talk but they communicate with other people but some time they are unable to understand each other talk it means more over the time they didn't understand what actually they want to say. So speech is the most important part of human being.

Let we take an example, suppose we are going to the doctor and we discuss our problem and if we are unable to understand our problem with the doctor then it will create a huge problem. It also happen doctor understand another problem and they start the treatment for another disease it creates a lots of problem. So in the end of the example speech is the most important part of human life. Human voice is of different types like utterance in the speech, prolongation in the word, pronunciation is not correct, etc. Human speech is most essential part of human life we can communicate with each other; we discuss many things without any problem. In the court room if the lawyer can't speak then how will he defend the case so it is very important to talk with

human being. If the lawyer wants to win the case and also if the person wants to win the case he will have to speak and try to defend herself/himself without speaking how the judge will understand the situation and makes the decision. Human voice is important because if we have any problem then we can say but if we don't have voice then how we can say and how I describe the situation? It is also possible anyone can't understand our problem or take wrong. So basically in this paper we differentiate the voice according to their age group. We recorded several voice of different persons and try to differentiate the voice according to the different age group with the help of fuzzy inference system. We use FLC (fuzzy logic control) system to differentiate the voice according to their age group.

II. FUZZY INFERENCE SYSTEM

A. *Fuzzy Logic*

Fuzzy set theory is also known as fuzzy logic which is used to compute the problem mathematically and this concept was established by Lofti Zadeh in 1965. It doesn't take the values either completely true or completely false. It helps in taking the values between the 0 & 1. It depends the value on "degree of truth". Fuzzy logic is a superset of Boolean (conventional) logic that handles the concept of partial truth, which is truth values between "completely true" and "completely false". Fuzzy logic is a multivalued logic and allows intermediate values to be defined. There is much

difference between fuzzy design and conventional design it means fuzzy logic design is easy as compared to conventional design.

In the conventional design there are five steps which we follow to simulte and design the whole process. These steps are:

1. Understanding the physical requirement

2. Develop a linear model

3. Simplifying the data

4. Develop an algorithm

5. Simulate implement and design

These are the five steps which we used to simulate the design in conventional design methodology. Conventional design methodology is difficult as compare to fuzzy logic design methodolgy. Conventional design methodology is the methodology which also helps in differentiating the human voice according to their age group. **(Ref Fiigure – 1)**

In the fuzzy logic design methodology there are only three steps to simulate the algorithm. In this methodology we apply if then rules to differentiate the voice according to their age. Steps include:

1. Understanding physical requirement

2. Designing using Fuzzy If-Then Rules

3. Simulate Implement and Design

These are the three steps which we used to differentiate the human voices according to their

different age group. In this methodology we use fuzzy if then rules except develop the model and simplifying the data in conventional methodology. We also skip to design the algorithm which we used to differentiate the voice we directly apply the fuzzy if-then rules and directly simulate the rules which easily helps in differentiating the voice according to their age group.

Fuzzy logic works on bivalued logic it means we take more than one input variable on which we can apply fuzzy if then rules on the fuzzy inference system and then we get the single output. **(Ref Figure – 2)**

In fig1 and fig2 it is clearly shown that fuzzy logic design methodology is easy as compared to conventional design methodology which helps in differentiating the human voice according to their age group. In fuzzy logic design we apply fuzzy if-then rules which are very easy to implement on the fuzzy inference system.

B. *Fuzzy Inference System*

Fuzzy Inference is the system which is used for mapping the input to an output using fuzzy logic. In the fuzzy inference system we use different membership functions, fuzzy if then rules, fuzzy logic operators to get the single output by taking more than one input variable.

Fuzzy inference process comprises of five parts:

1. Fuzzification of the input variables

2. Application of the fuzzy operator (AND or OR) in the antecedent

3. Implication from the antecedent to the consequent

4. Aggregation of the consequents across the rules

5. Defuzzification

These are the five parts which includes in the process of fuzzy inference system to get output. There are many fields on which we apply fuzzy inference system such as automatic control, data classification, decision analysis, expert systems, and computer vision.

Because of its versatile nature, the fuzzy inference system is known by a various names like:

1. Fuzzy-Rule-Based Systems
2. Fuzzy Expert Systems
3. Fuzzy model
4. Fuzzy associative memory
5. Fuzzy logic controller

These are the various names of fuzzy inference system due to its versatile nature. **(Ref Figure – 3)**

C. *Fuzzy If-Then Rules*

Fuzzy If-Then rules are the statement which is used to formulate the conditional statement that comprises the fuzzy logic [1].

A single fuzzy if then rules assumes the form

$$\text{If } x \text{ is } A_1 \text{ Then } y \text{ is } B_2$$

where A and B are linguistic values defined by fuzzy sets on the ranges (universes of discourse) X and Y, respectively. The if-part of the rule "x is A" is called the *antecedent* , while the then-part of the rule "y is B" is called the *consequent*.

In other words we can write this statement in mathematical form

$$\text{If } A_1 \text{ Then } B_2 \text{ or } A_1 \rightarrow B_2$$

Let take an example there is a human speech system of different age groups which is controlled by fuzzy inference system. The system decides which voice belongs to which age group. The fuzzy engine periodically compares the human speech dataset and the target age group, and produces a command to declare the final group of the human speech. **(Ref Figure – 4)**

In the above example we have to decide the human speech with the help of fuzzy logic system according to the target age group by making some fuzzy if then rules. But before making the rules we have to decide the linguistic variables of human speech on which human speech varies. In this work we decide the linguistic variables like Very High (VH), High (H), Average (A), Low (L), and Very Low (VL). On the basis of these linguistic variables we decide which speech belongs to which age group.

These are some rules which we used to decide which speech belongs to which age group. These rules are making on the basis of the input variables.

We can make any number of fuzzy if-then rules for any system it depends on the researcher.

In this work we take three parameters like Maximum Frequency (MF), Standard Deviation (SD) and Covariance (CV) of the speech.

Fuzzy Rules for this system-

1. If MF is VH and SD is VH and CV is VH then OUTPUT is VH

2. If MF is VH and SD is VH and CV is H then OUTPUT is VH

3. If MF is VH and SD is VH and CV is A then OUTPUT is VH

4. If MF is VH and SD is VH and CV is L then OUTPUT is H

5. If MF is VH and SD is VH and CV is VL then OUTPUT is H

6. If MF is VH and SD is H and CV is VH then OUTPUT is VH

7. If MF is VH and SD is H and CV is H then OUTPUT is VH

8. If MF is VH and SD is H and CV is A then OUTPUT is H

9. If MF is VH and SD is H and CV is L then OUTPUT is H

10. If MF is VH and SD is H and CV is VL then OUTPUT is H

III. PARAMETERS

There are many parameters which help in deciding the human voice according to their age group. Parameters like pitch, zero crossing rate,

short term energy, total harmonic distortion, power amplitude, maximum frequency, standard deviation, covariance etc. On the basis of these parameters we can distinguish the voice according to their gender and their age group. In this research we take three parameters to distinguish the human voice according to their age.

A. ***Frequency***- Voice frequency or voice bands are one of the frequencies, within part of the audio range, which is being used for the speech. Frequency of a periodic wave can be defined as the number of times a repeating pattern occurs in a period of time.

Example: In an electromagnetic wave how many no of crest and trough cycles per second. Human voice s plotted on the graph with time and its amplitude, time is plotted on x-axis and amplitude on y-axis.

According to the telephony, the usable voice frequency band ranges from 300 Hz to 3400 Hz approximately [3]. This is the reason that the ultra-low frequency band of the electromagnetic spectrum between 300Hz to 3000 Hz [3] is also referred to as voice frequency. Frequency is of three types in the voice like maximum frequency, medium frequency, and mean frequency. We take maximum frequency (MF) as one of the parameter of the voice for voice recognition as

per their age group. Frequency of each group is different from *other age groups*.

In the below image it is defined that the parameter maximum frequency of the speech is shown with five linguistic variables. Now we are working on this parameter with the help of fuzzy logic. We calculate the values of maximum frequency of the human speech with the help of oscilloscope. **(Ref Figure – 4 (b))**

B. ***Standard Deviation (SD)-*** The concept of standard deviation was introduced by Karl Pearson in 1983. It is the measure of dispersion of the set of data from its mean. Standard deviation helps in measure the absolute variability of a distribution; the higher the dispersion or variability, the greater is the standard deviation and greater will be the magnitude of the deviation of the value from their mean [4].

Standard Deviation is a quantity which is derived from distribution of marks from the regularizing collecting samples. It is often used in norm- referenced tests to diagnose language impairment.

In other words, standard deviation is a statistic that measures the dispersion of the various dataset relative to its mean and is calculated as the square root of the variance. Standard deviation is calculated as the square root of the

variance by determining the variation between each data point relative to the mean [5]. We take standard deviation as the parameter for the recognition of the voice and we can implement in matlab with the help of fuzzy logic toolbox as we can see in fig5 with five different linguistic variables as we describe in first parameter maximum frequency. In this work standard deviation explains that the difference between the maximum frequency and the mean frequency. It means that after how many times maximum frequency repeats.

We can calculate standard deviation with the help of given formula –

$$\sigma = \sqrt{\frac{\sum[F_{max} - \overline{F_0}]}{n}}$$

Where, F_{max} defines the maximum frequency,

F_0 defines the mean frequency and

n is the no of frequencies we calculated per speech. **(Ref Figure – 5)**

C. Coefficient Of Variation (CV) -

This is another parameter which we take for the recognition of the human speech as per their age group. Covariance is also known as coefficient of variation and it is calculated as the ratio of standard deviation to the mean frequency. It is a standardized measure of dispersion of a probability distribution or frequency distribution.

$$CV = SD/Mean$$

The coefficient of variation also gives us the general "feeling" about the performance of the method. If the value of the coefficient of variation is less it means the performance of the system is good otherwise the performance of the system is not too good. The performance of the system depends on the value of the standard deviation which we calculate on the basis of the frequencies of the voice which we taken. **(Ref Figure – 6)**

D. Final Age Group-

This is the output of this work, in this it defines the age group according to the voice which we collect from the different peoples. We take five age groups:-

1. Infancy
2. Childhood
3. Adolescence
4. Adulthood
5. Senior

E. Oscilloscope-

Oscilloscope is the machine which is used for calculating the frequencies of the voice with the help of the mic or mike. We use oscilloscope for calculating the frequencies of the different voice then we calculate the standard deviation. After calculating these values we calculate the value of coefficient of variation with the help of the formula. In this work till now we calculate the values of these parameters.

Previously we called oscilloscope as oscillograph and informally known as CRO (Cathode Ray Oscilloscope) or DSO (for the more modern digital storage oscilloscope). , is a type of electronic test instrument that graphically displays varying signal voltages, usually as a two-dimensional plot of one or more signals as a function of time. The waveform can then be analyzed for properties such as amplitude, frequency, rise time, time interval, distortion, and others. Modern digital instruments may calculate and display these properties directly. Originally, calculation of these values required manually measuring the waveform against the scales built into the screen of the instrument [7]. **(Ref Figure – 7, 8, 9)**

IV. Results

In this research we calculate the values of three parameters using oscilloscope. We take several voices of different age group the with the help of mike we record the voice then pass that voice signal in to the oscilloscope then we calculate the frequencies of different voices and then standard deviation then coefficient of variation. **(Ref Table – 1)**

CONCLUSION

In this work we calculate the values of three parameters of different voices of different age group with the help of oscilloscope. With the help

of these values we decide that which voice belongs to which age group. The objective of this work is to find out that the voice which we taken is of which age group. Till now there is no work is done on the old age people's voice. So in this work we also calculate the values of old age people along with the teenagers.

ACKNOWLEDGMENT

REFERENCES

[1]https://edoras.sdsu.edu/doc/matlab/toolbox/fuzzy/fuzzytu5.html

[2]http://web.abo.fi

[3]https://en.wikipedia.org/wiki/Voice_frequency

[4]https://economictimes.indiatimes.com/definition/standard-deviation

[5]https://www.investopedia.com/terms/s/standarddeviation.asp

[6] https://en.wikipedia.org/wiki/Oscilloscope

[7] Kularatna, Nihal (2003), "Fundamentals of Oscilloscopes", *Digital and Analogue*

Instrumentation: Testing and Measurement, Institution of Engineering and Technology, pp. 165–208

[8] Shivaji J Chaudhari, Ramesh M Kagalkar, "Methodology for Gender Identification, Classification and Recognition of Human Age", International Journal of Computer Applications (0975-8887), National Conference on Advances in Computing, 2015

[9]Arul Valiyavalappil Haridas, Ramalatha Marimuthu, Vaazi Gangadharan Sivakumar, "A Critical review and analysis on techniques of Speech recognition: The road ahead", International Journal Of Knowledge-based and Intelligent Engineering System, 2018.

[10] Syed Mostafa Mmirhassani, Alireza Zourmand, and Hua-Nong Ting, "Age estimation based on children's voice: A fuzzy-based decision fusion strategy", The Scientific World Journal, Vol 2014, Article ID 534064

LIST OF FIGURES

Figure- 1: Conventional Design Methodology
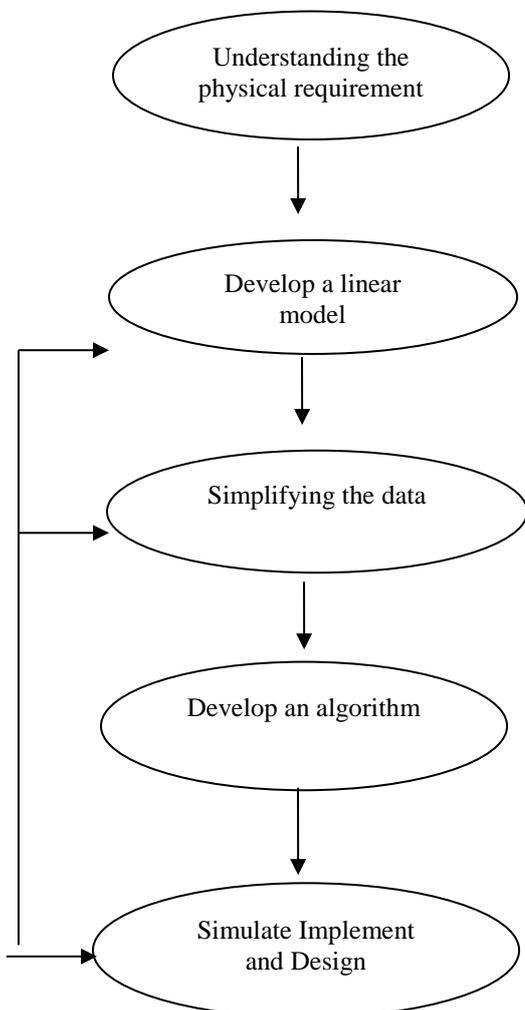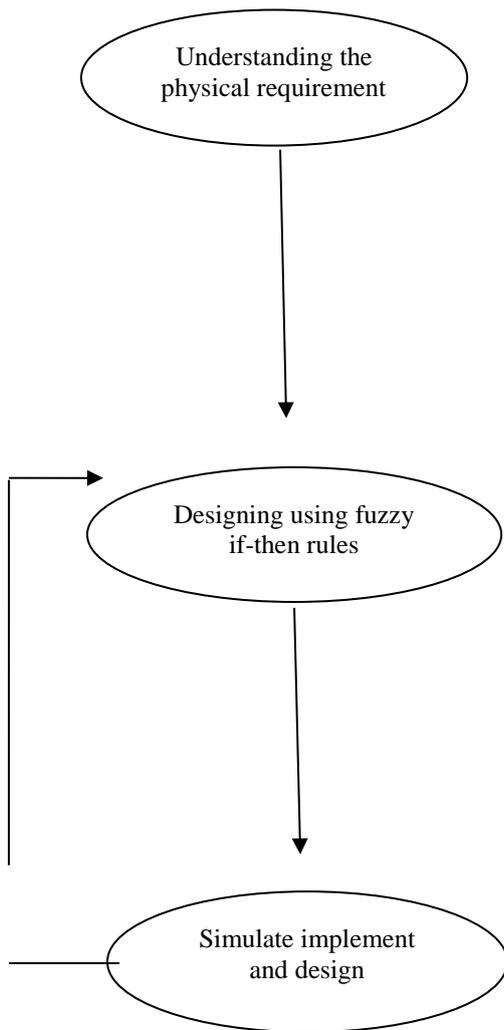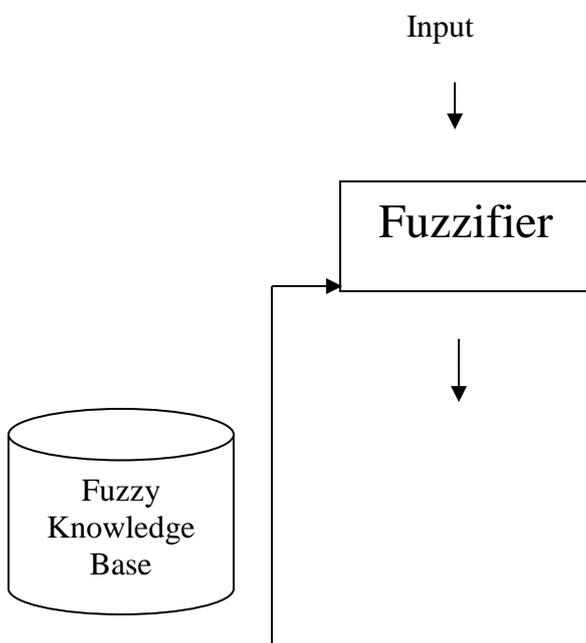
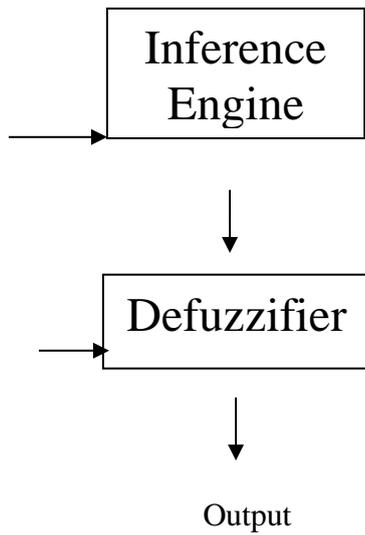**Figure - 2: Fuzzy Logic Design Methodology**

Figure - 3: Architecture of Fuzzy inference system

Inference Engine

Defuzzifier

Output

Figure - 4: Simple FLS to decide the Human Speech according to their Age Group

Target

Age Group   Command

Infancy/childhood/

Adolescense/ Adulthood

Senior Citizen

FLS   Human Speech   Age Group

Human Speech
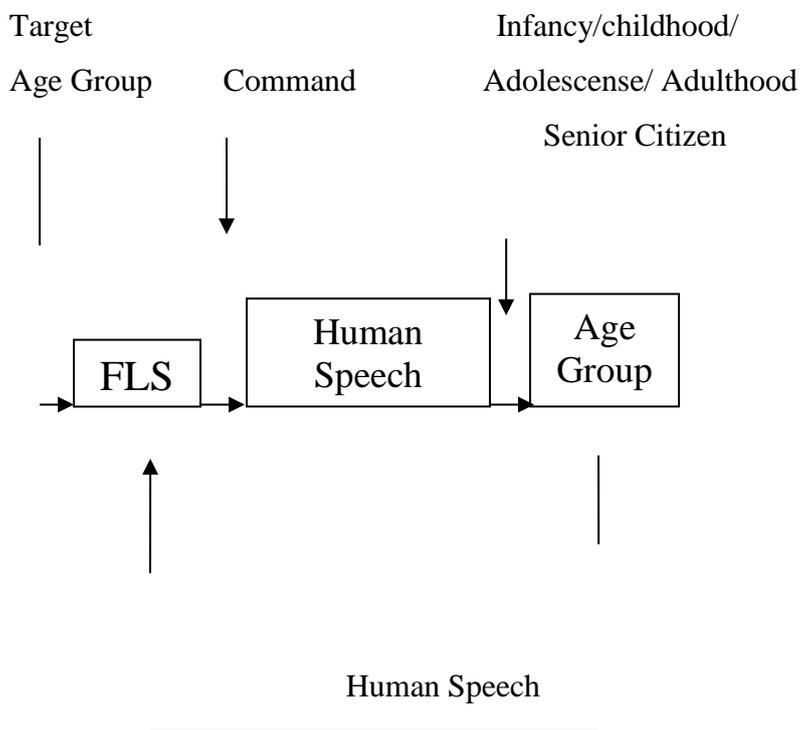
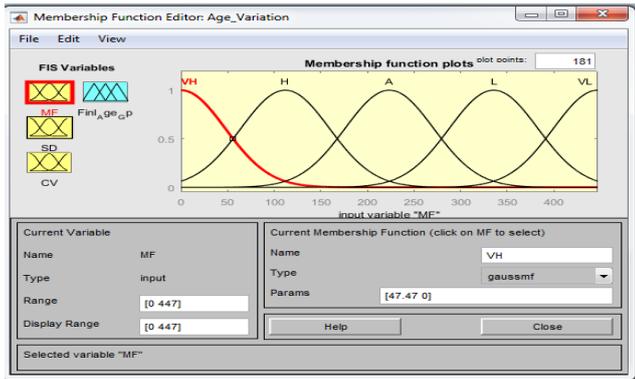**Figure- 4 (b): FIS for input Maximum Frequency**

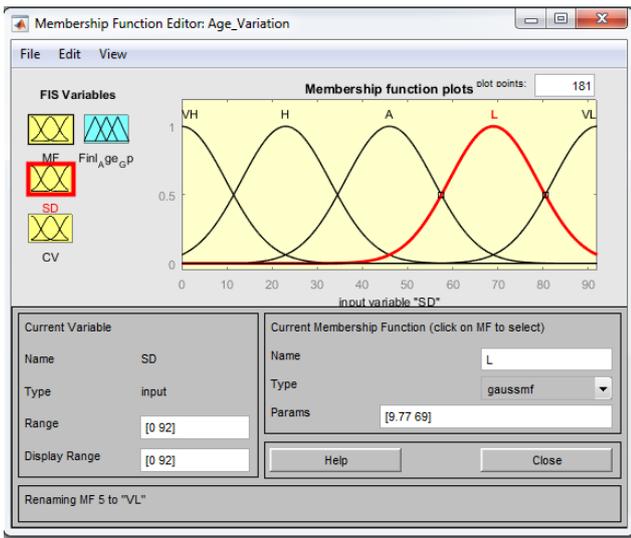**Figure - 5: FIS for input Standard Deviation**

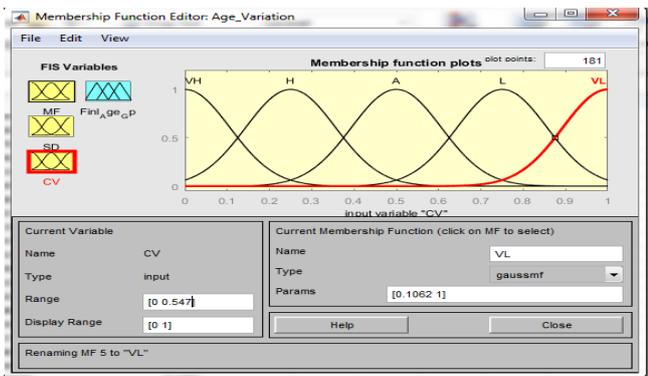**Figure - 6: FIS for input Coefficient of variation**
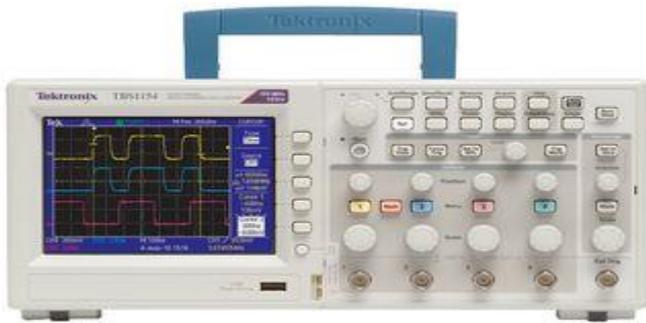
Figure - 7: Oscilloscope Machine

Figure- 8: Oscilloscope Machine Screen while passing a voice signal
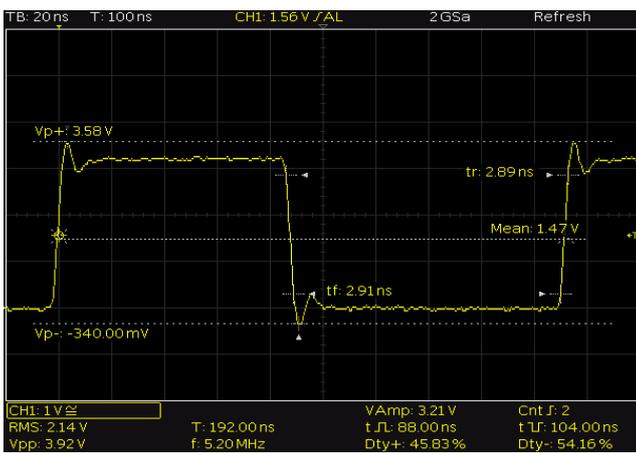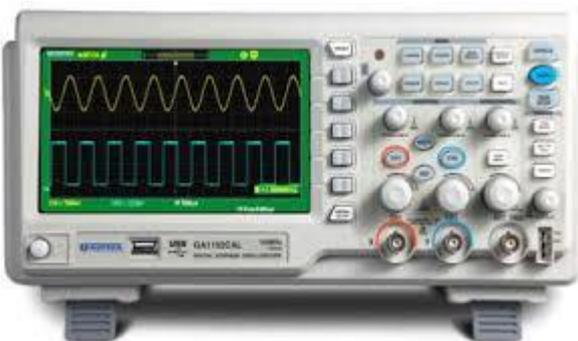

Figure- 9: Oscilloscope Machine Screen to show the difference between analog and digital image

LIST OF TABLE:-

Table 1: Values of different age group

Age	Maximum Frequency	Standard Deviation	Coefficient of variation
1	139.125	26.56484212	0.113
1	141.235	16.54889151	0.0822
1	142.256	19.25456745	0.155
2	123.235	12.3854312	0.113
2	334.235	34.12354	0.146
3	316.152	39.8956323	0.173
3	345.235	40.87946312	0.161
3	344.785	41.45874531	0.181
6	419.901	46.0247914	0.152
6	406.985	28.0029507	0.106
6	434.968	47.85565476	0.157
8	446.395	64.16707519	0.185
8	335.977	20.10317766	0.075
11	136.235	9.536214258	0.084
11	276.191	21.87545068	0.109
12	246.462	11.21453215	0.062
13	266.437	17.92815389	0.094
13	261.463	27.89381784	0.149
15	295.279	46.25677615	0.201
15	297.217	24.02863379	0.098
15	292.81	49.19990815	0.218
21	208.146	18.31041018	0.129
21	212.858	17.07614957	0.116
21	238.368	21.16344014	0.142

28	214.387	24.98446775	0.238
28	122.85	7.142365709	0.075
28	121.663	6.43741713	0.071
32	149.176	14.30150646	0.14
32	119.009	7.930665654	0.086
35	213.508	14.10281899	0.094
38	256.491	28.19998044	0.176
38	229.742	17.57457687	0.116
42	147.334	15.9908743	0.155
42	173.237	22.96178941	0.21
42	246.352	29.25714054	0.19
46	237.316	21.90480672	0.138
46	219.167	14.63395804	0.118
46	146.004	9.48854045	0.079
66	256.139	25.41665046	0.182
66	233.145	21.99037525	0.158
66	286.774	26.95989921	0.233
80	148.235	26.23145645	0.264
80	145.248	22.12354659	0.24
83	165.234	26.13564655	0.266
90	203.654	23.13514531	0.316
90	217.235	29.32165479	0.379
90	244.236	23.46584532	0.288