International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 12 No. 1 Jul. 2015, pp. 33-61

http://www.ijias.issr-journals.org/

HUMAN VOICE ACTIVITY DETECTION USING WAVELET

Md. Shahadat Hossain, Ariful Islam, and Dr. Md. Rafiqul Islam

Mathematics Discipline, Science Engineering and Technology School, Khulna University, Khulna-9208

ABSTRACT: Wavelet has wide range of use in the present scientific universe. At present using wavelet through MATLAB different types of tasks are done. For instance biometric recognition (fingerprint recognition, voice recognition, iris recognition, face recognition, pattern recognition and signature recognition), signal processing, human voice activity detection etc. are done using wavelet and wavelet transform. Among these here I have discussed about "Human Voice Activity Detection". At first a human voice is taken as the input sound to MATLAB command window using a good headphone for a few second. Then the sound taken as input give a graphical representation that is saved for future activities. After that using the wavelet toolbox of MATLAB the image of the input sound is taken for analyzing it. Using discrete wavelet transform the image is analyzed. During this analysis a "10 level wavelet" tree is generated by Haar wavelet with 10 decomposition level. At the same time the original signal is reconstructed. At the first time six different human voice activities of the same persons are analyzed. The Norm and the SNR (Signal to Noise Ratio) are counted. The data of the SNR are counted in decibel (db.) unit. Also the bit rates of the three different voice are counted. In this way total 18 different experiments are done for the different five persons where except the first person for all the person three experiments are dine.. The numerical data of the experiments are shown as graphical representation as well as in histogram analysis. In this process the whole experiments are done for the activity detection of human voice.

KEYWORDS: Wavelet, SNR, Bit rate, Human voice, Histogram.

1 INTRODUCTION

Recently, human-machine interface system based on speech attracts much interest, supporting with the rapid improvement of the CPU performance. The speech-based interface is greatly based on speech recognition, in which the information of voice activity segments (VAS) is effective to improve the recognition rate. For the voice activity detection, various methods have been proposed. They use the features of speech signal, such as transition of the power [1], harmonic structure in spectrum [2] [11] [3] and the existence of signal source directionality [4]. In these methods, acquired speech is usually assumed to be sufficiently clean, due to the preprocessing used in speech recognition and compression for transmission. However at indoor environments where the interface is ordinarily used, there are various localized interferences arriving from particular direction such as the sound of closing door, etc. For these non-stationary interferences, the conventional methods do not realize sufficient performance, because of stationarity and whiteness assumption to noise. Kaneda [10] [12] proposed an effective VAD method available for these non-stationary interferences, using their high performance speech emphasizing system"AMNOR (Adaptive Microphone array for Noise Reduction)". He uses microphone array to discriminate signals utilizing direction difference between speech and interference. However, target speech and interference are required to arrive from sufficiently separated direction due to the spatial resolution in AMNOR. This limitation critically restricts the applicable condition of the method. In this research, we propose a new method to be robust to the direction of interference, with microphone array signal processing in the wavelet domain to integrate the time, frequency and spatial information of speech signal.

Corresponding Author: Md. Shahadat Hossain

1.1 VOICE ACTIVITY DETECTION (VAD)

Voice activity detection (VAD) refers to the problem of distinguishing speech from non-speech regions (segments) in an audio stream. The non-speech regions could include silence, noise, or a variety of other acoustic signals. VAD is challenging in low signal-to-noise ratio (SNR), especially in non-stationary noise, because both low SNR and a non-stationary noisy environment tend to cause significant detection errors. There is a wide range of applications for VAD, including mobile communication services [5], real-time speech transmission on the Internet [6], noise reduction for digital hearing aid devices [7], automatic speech recognition [8], and variable rate speech coding [9]. Voice activity detection (VAD), also known as speech activity detection or speech detection**,** is a technique used in speech processing in which we detect the low bit rate speech and high bit rate speech, also can distinguish between high vocal speech and low vocal speech of human. The main uses of VAD are in speech coding and speech recognition. It can facilitate speech processing, and can also be used to deactivate some processes during non-speech section of an audio session. VAD is an important enabling technology for a variety of speech-based applications. Therefore various VAD algorithms have been developed that provide varying features and compromises between latency, sensitivity, accuracy and computational cost. Some VAD algorithms also provide further analysis, for example whether the speech is voiced, unvoiced or sustained. Voice activity detection is usually language independent. There are many voice detection technique already exists like CMU Sphinx, Julius, kaidi, Bing, SILVIA, Vlingo, Microsoft Tellme, Ask Ziggy, wavelet etc. Among these Technique wavelet is used and compressed signal by wavelet technique and which gives better results for lossless compression. The practical implementation of voice signal compression schemes is very similar to that of sub band coding schemes. As is case sub band coding, we compress the signal (analysis) using different wavelets. The output of the compression is dawn sampled and comparison among the compression signal. Wavelet analysis can be used to divide the information of a signal into approximation and detail sub signals shows the vertical and horizontal details or changes in the signal. If these details are very small then they can be set to zero without significantly zero knows as threshold. The greater the number of zeros the greater the compression ratio. The amount of information retained by a signal after compression and decompression is known as the retained energy and this is proportional to the sum of square of the matrix values. If the energy retained 100% then the compression is known as lossless as the signal can be reconstructed exactly. This occurs when the threshold value is set to zero, meaning that the detail has not been changed. Ideally, during compression the number of zeros are obtained more energy retention will be as high as possible. We know wavelet packet to perform significantly better than wavelets for compression of signals with large amount texture and it is also point out the perceived signal quality is significantly improved using wavelet packets instead of wavelets especially in the textured regions of the signals. This chapter deals with speech compression based on discrete wavelet transforms. We used English words (only hello) for this experiment. We have successfully compressed and reconstructed the words with perfect audibility by using wavelet technique. Speech compression is the technology of converting human speech into an efficiently encoded representation that can later be decoded to produce a close approximation of the original signal. The wavelet transform of a signal decomposes the original signal into wavelets coefficients at different scales and positions. These coefficients represent the signal in the wavelet domain and all data operations can be performed using the corresponding wavelet coefficients.

In our study we obtain code form wavelet coding and then the code is simulated using MATLAB. From the results we noticed that the performance of Wavelet Coding which can detect the distinguish between low bit rate speech and high bit rate speech; also can distinguish between high vocal speech and low vocal speech. Four men (A, B, C, D, E, male) participate in this experiment with different age and voice. Here we take 18 (Eighteen) experimental voice via headphone and for the resultant discursion we calculate only $\mathbf{L_1}$, $\mathbf{L_2}$ Norm, Therehold value and SNR. All of experiments are given in below.

1.2 WHAT IS VOICE RECOGNITION, AND WHY IS IT USEFUL IN A VIRTUAL ENVIRONMENT?

Voice recognition is "the technology by which sounds, words or phrases spoken by humans are converted into electrical signals, and these signals are transformed into coding patterns to which meaning has been assigned". While the concept could more generally be called "sound recognition", we focus here on the human voice because we most often and most naturally use our voices to communicate our ideas to others in our immediate surroundings. In the context of a virtual environment, the user would presumably gain the greatest feeling of immersion, or being part of the simulation, if they could use their most common form of communication, the voice. The difficulty in using voice as an input to a computer simulation lies in the fundamental differences between human speech and the more traditional forms of computer input. While computer programs are commonly designed to produce a precise and well-defined response upon receiving the proper (and equally precise) input, the human voice and spoken words are anything but precise. Each human voice is different, and identical words can have different meanings if spoken with different inflections or in different contexts. Several approaches have been tried, with varying degrees of success, to overcome these difficulties.

1.3 L_p NORM

For finite p , L_p Norm in $c[a,b]$ is defined as

$$\|f\|_p = \left[\int_a^b |f(x)|^p dx\right]^{\frac{1}{p}} \; ; 1 \leq p < \propto$$

For discreate function it can be defined as

$$\|f\|_p = \left[\sum_{i=1}^{n} |f(x_i)|^p\right]^{\frac{1}{p}}$$

Where $\{x_i\}$ are the components of f .

If we put $p = 1$ in the above equation then $\|f\|_1$ is called $\mathbf{L_1}$ **Norm**.

If we put $p = 2$ in the above equation then $\|f\|_2$ is called $\mathbf{L_2}$ **Norm**.

1.4 SIGNAL TO NOISE RATIO (SNR)

This value gives the quality of reconstructed signal. Higher the value, the better:

$$SNR = 10log_{10}\frac{\sigma x^2}{\sigma e^2}$$

σx^2 Is the mean square of the speech signal and σe^2 is the mean square difference between the origiand reconstructed signals.

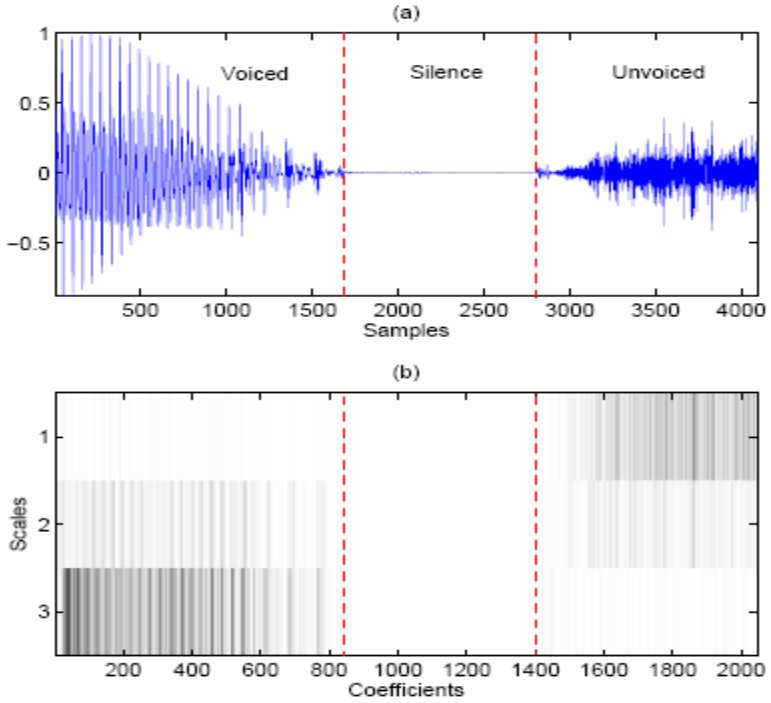
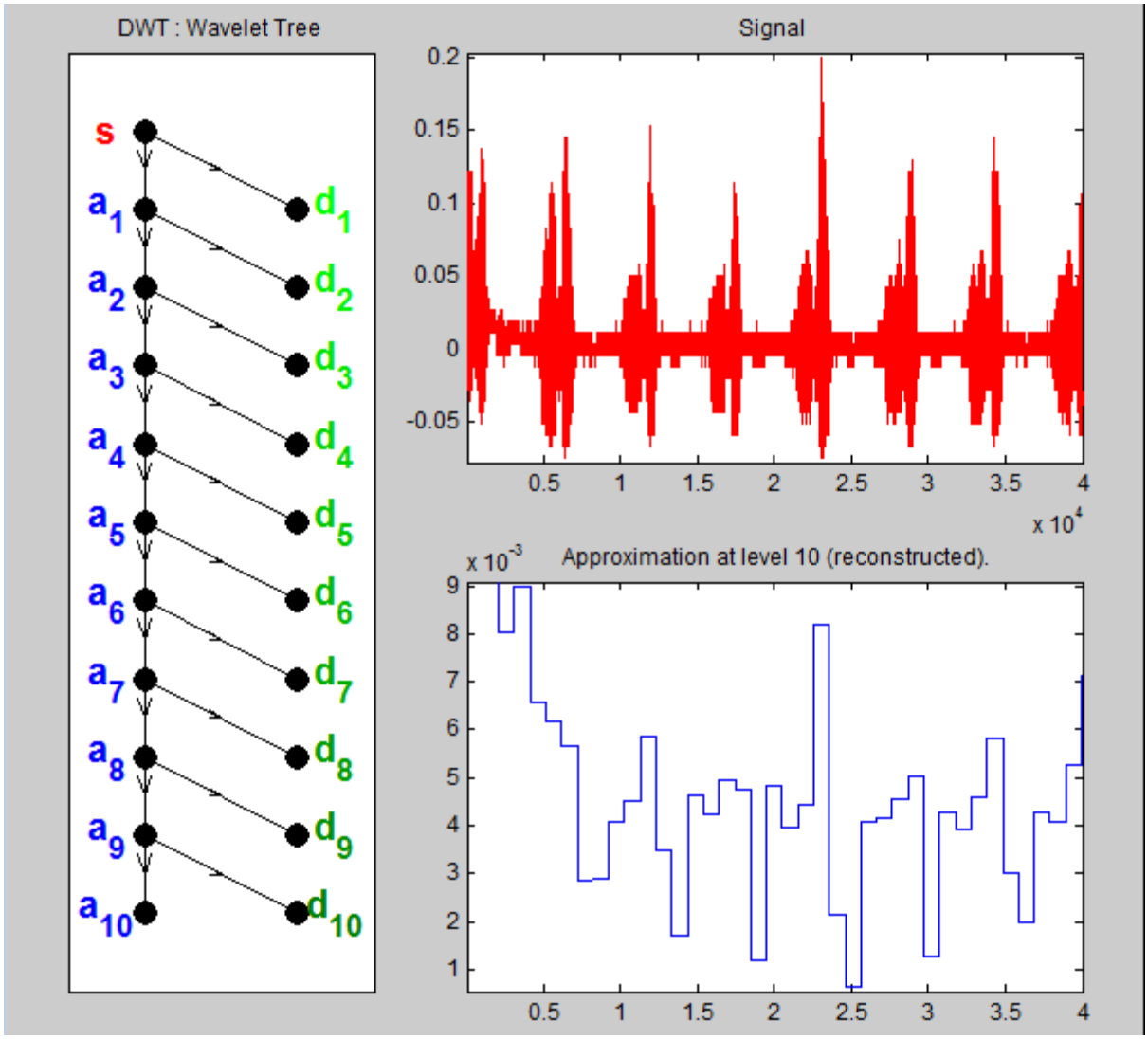


Fig. 1. (a) speech segment consisting of voiced, unvoiced and silence frames, (b) Power variation of detail coefficients

1.4.1 EXPERIMENT 1: ORIGINAL SPEECH SIGNAL OF MR. "A"

CODING OF SPEECH SIGNAL:

Mr. "A" is talking with headphone: Hello, Hello, Hello, Hello, Hello, Hello, and Hello (7 Times)



X axis→ time (ms), Y axis→

Fig. 2. Original speech signal of Mr."A" with 10 level decomposition and decomposition tree.

$\mathbf{L_1\ Norm = 481.8}$

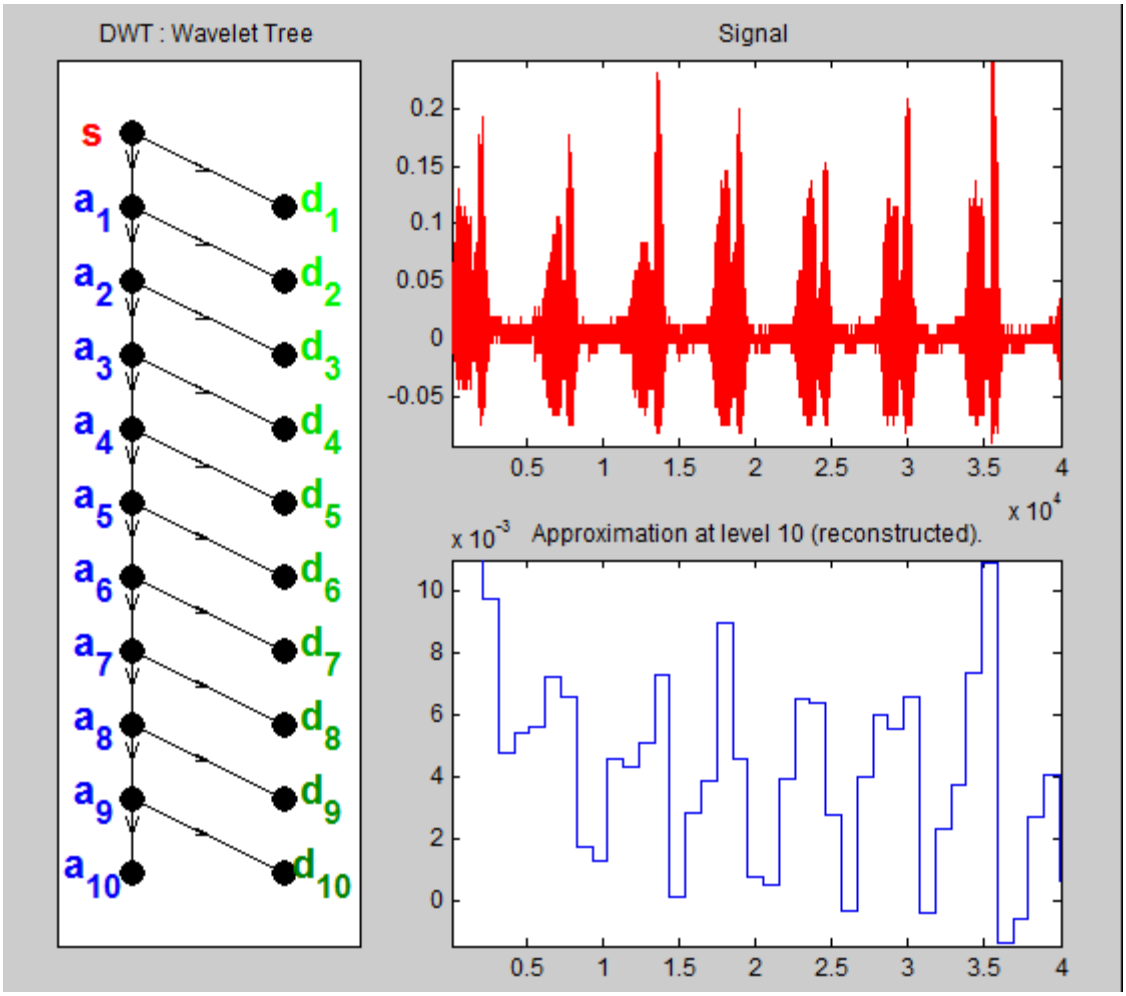
$\mathbf{L_2\ Norm = 3.845}$

Signal to noise ratio = 6.81 db

1.4.2 EXPERIMENT 2: ORIGINAL SPEECH SIGNAL OF MR. "A"

CODING OF SPEECH SIGNAL:

Mr. "A" is talking with headphone: Hello, Hello, Hello, Hello, Hello, Hello, and Hello (7 Times)



X axis→ time (ms), Y axis→

Fig. 3. Original speech signal of Mr."A" with 10 level decomposition and decomposition tree.

L_1 Norm $= 587.5$

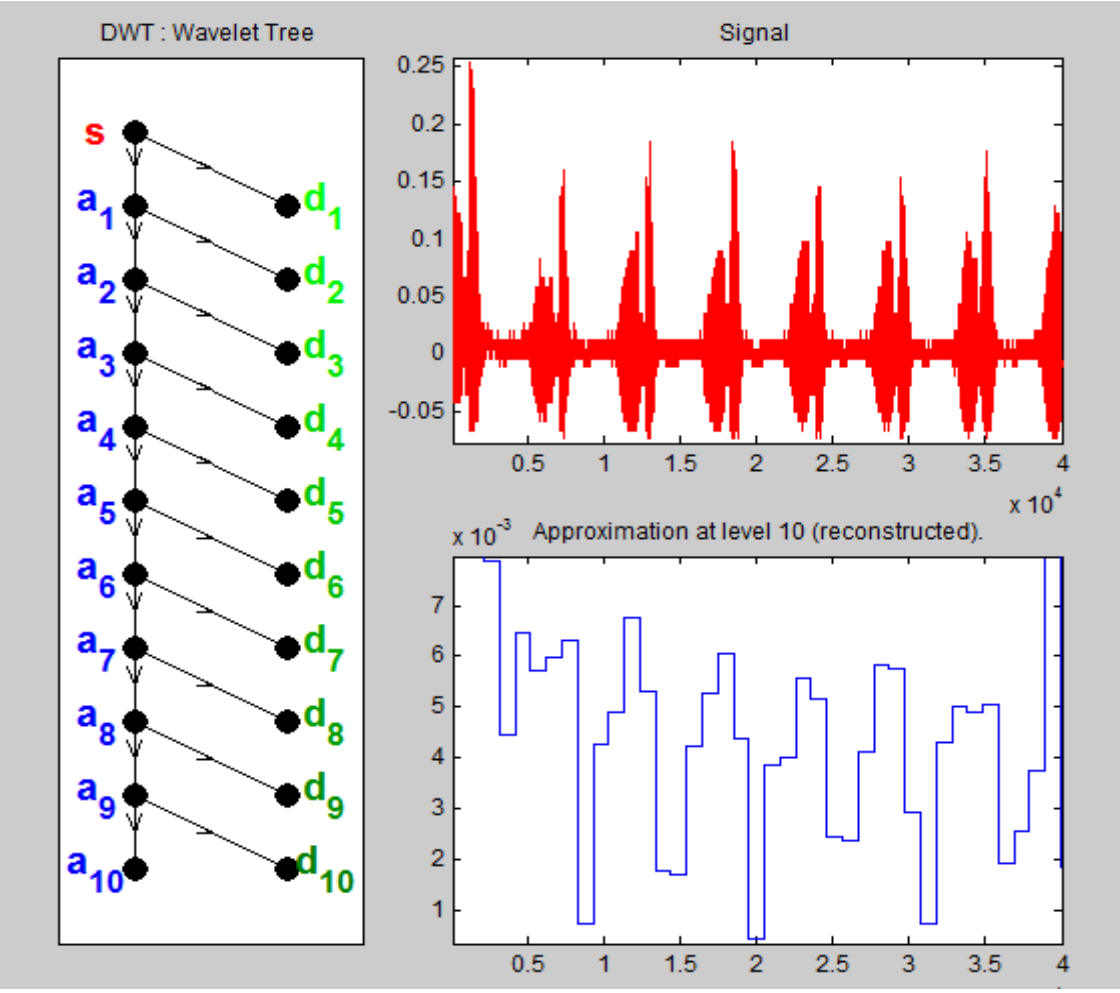
L_2 Norm $= 5.058$

Signal to noise ratio = 6.00 db

1.4.3 EXPERIMENT 3: ORIGINAL SPEECH SIGNAL OF MR. "A":

CODING OF SPEECH SIGNAL:

Mr. "A" is talking with headphone: Hello, Hello, Hello, Hello, Hello, Hello, Hello, and Hello (8 Times)



X axis→ time (ms), Y axis→

Fig. 4. Original speech signal of Mr."A" with 10 level decomposition and decomposition tree.

$\mathbf{L_1\ Norm = 570.1}$

$\mathbf{L_2\ Norm = 4.78}$

Signal to noise ratio = 5.87 db

1.4.4 DATA CHART FOR MR. "A":

	1st voice	2nd voice	3rd voice
$\mathbf{L_1}$ **Norm**	481.8	587.5	570.1
$\mathbf{L_2}$ **Norm**	3.84	5.05	4.78
SNR	6.81	6.00	5.81

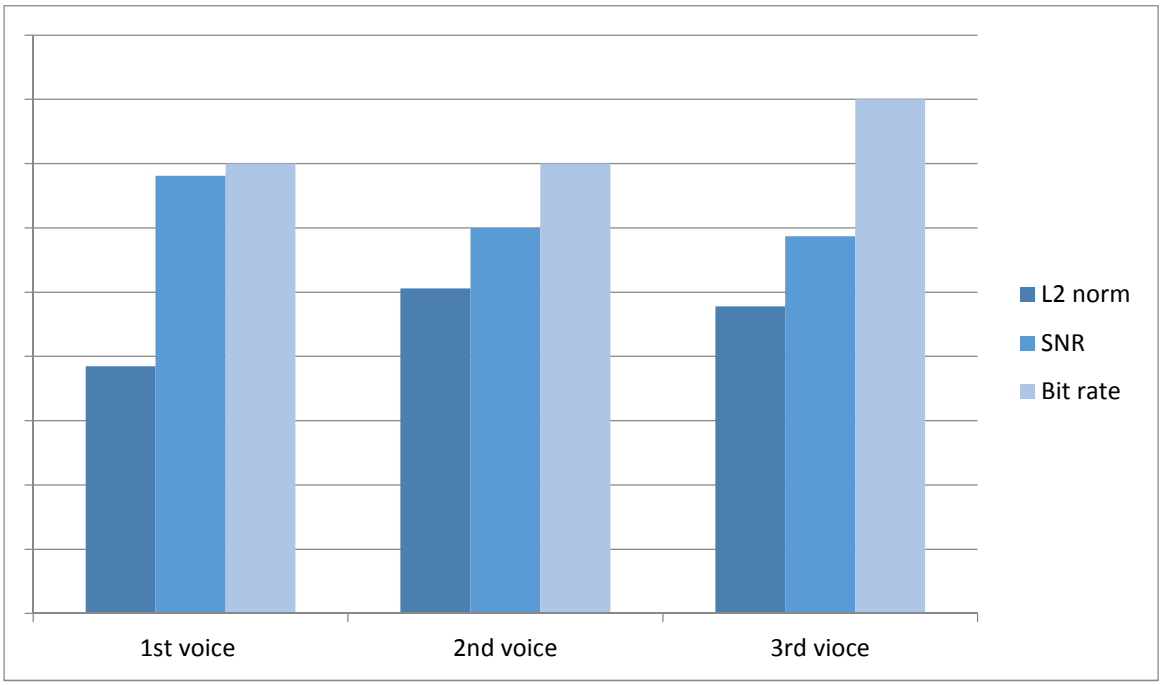


Fig. 5. Three experimental voice data chart of Mr."A".

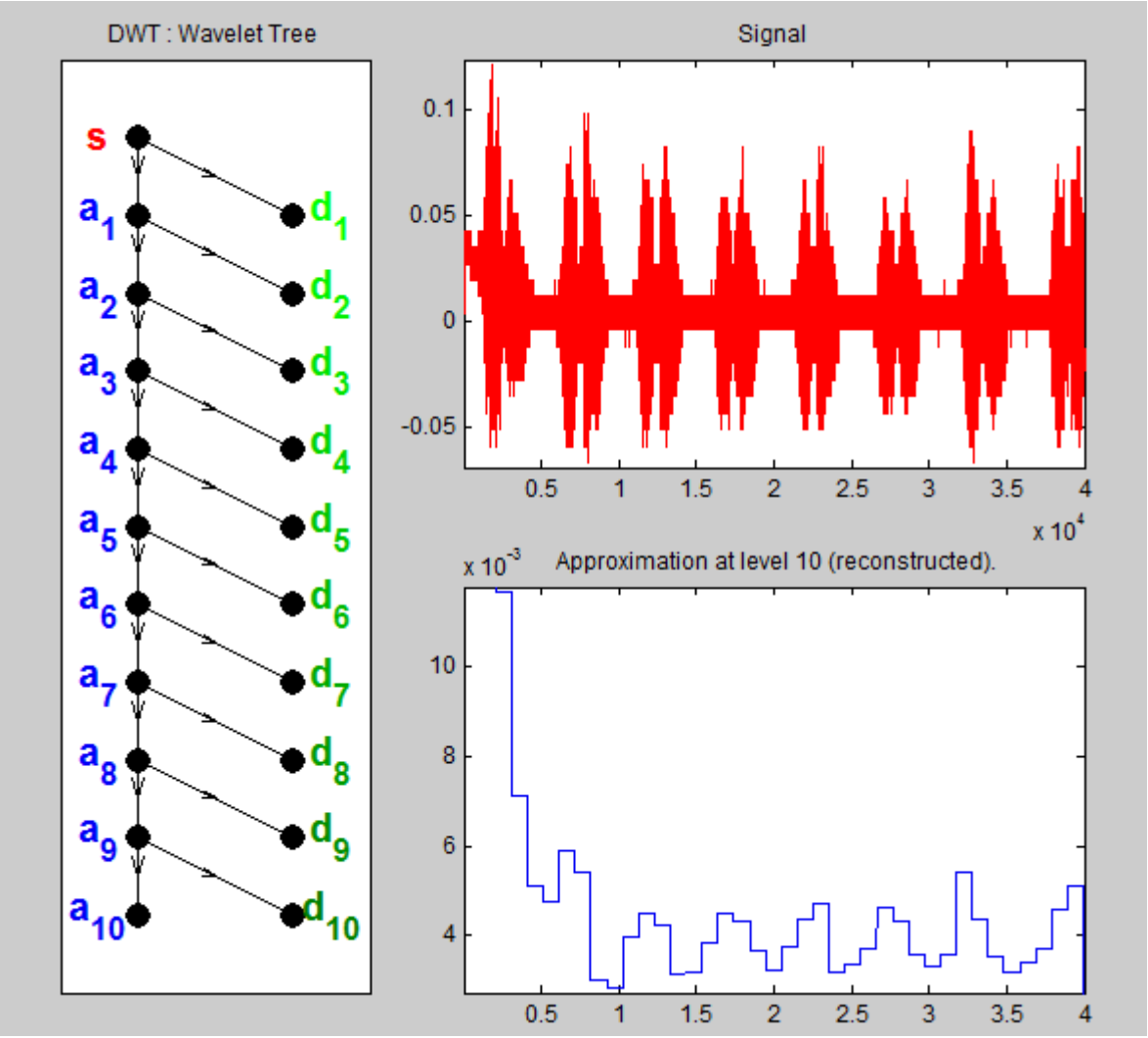
Summary:

From the above chart it is concluded that the SNR value is minimum for the high bit rate voice (fast voice) than that the slow bit rate voice (slow voice) within the same decibel (db) value. If we increase the volume of vocal chord in different cases then the result will be change. With the increase of volume of vocal chord, the value of L1 norm and L2 norm is increase respectively. For a same range bit rate voices with different volume, SNR value may be increased or decreased but L1and L2 norm must increased with the increase of volume. By this way we can say that the 2nd speech has the height volume i.e. height db value among the 3 speech and the 3rd Speech is faster voice than the other two.

1.4.5 EXPERIMENT 4: ORIGINAL SPEECH SIGNAL OF MR. "A":

CODING OF SPEECH SIGNAL:

Mr. "A" is talking with headphone: Hello, Hello, Hello, Hello, Hello, Hello, Hello, and Hello (8 Times)



X axis→ time (ms), Y axis→

Fig. 6. Original speech signal of Mr. "A" with 10 level decomposition and decomposition tree.

$\mathbf{L_1\ Norm = 491.2}$

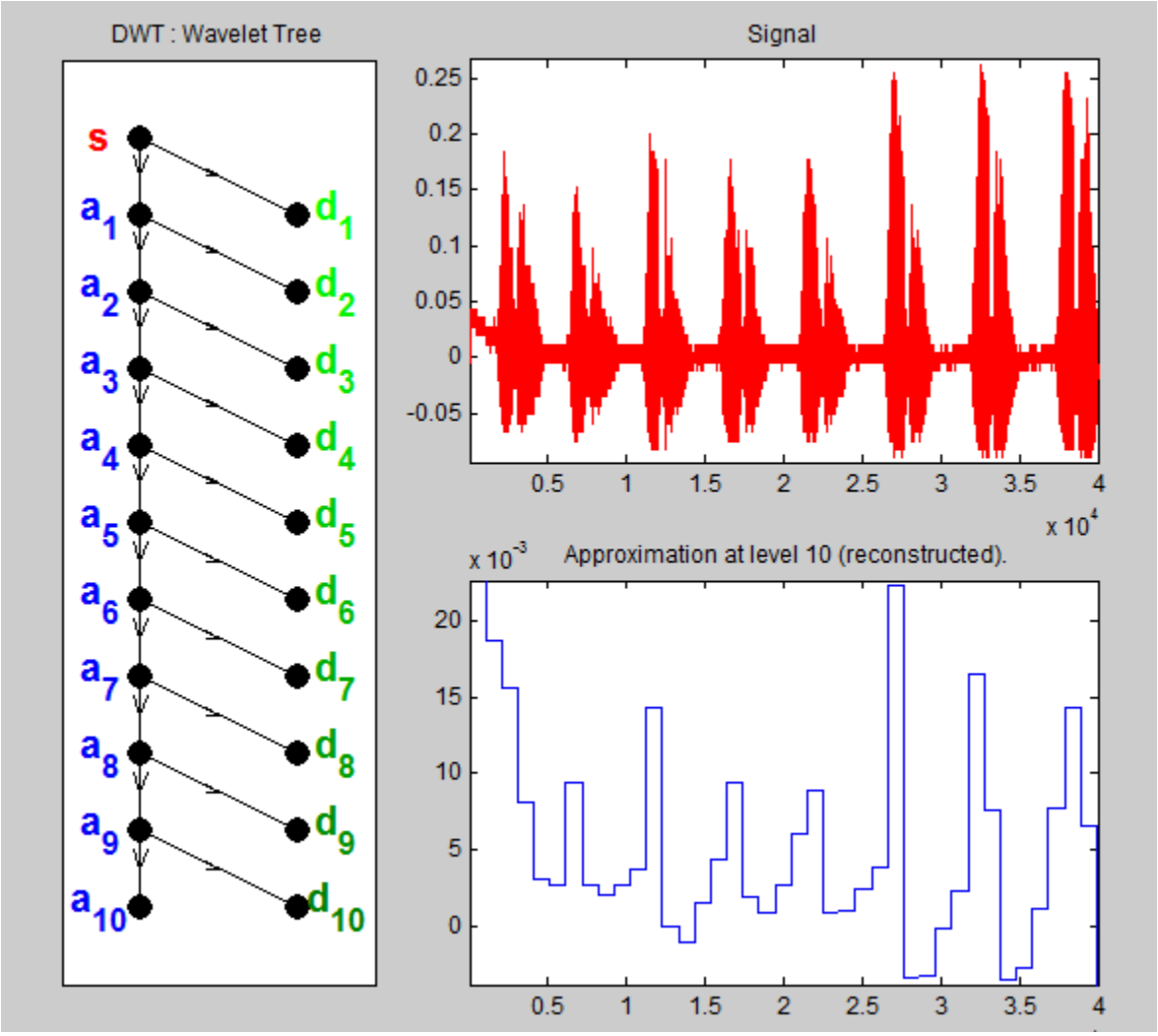
$\mathbf{L_2\ Norm = 3.51}$, *Signal to noise ratio = 5.83 db*

1.4.6 EXPERIMENT 5: ORIGINAL SPEECH SIGNAL OF MR. "A":

CODING OF SPEECH SIGNAL:

Mr. "A" is talking with headphone: Hello, Hello, Hello, Hello, Hello, Hello, Hello, Hello

(8 Times with loud enough)



X axis→ time (ms), Y axis→

Fig. 7. Original speech signal of Mr."A" with 10 level decomposition and decomposition tree.

L_1 Norm = 842.8

L_2 Norm = 7.00

Signal to noise ratio = 5.87 db

1.4.7 EXPERIMENT 6: ORIGINAL SPEECH SIGNAL OF MR. "A":

CODING OF SPEECH SIGNAL:

Mr. "A" is talking with headphone: Hello, Hello, Hello, Hello, Hello, Hello, Hello,

(7 Times with loudly)

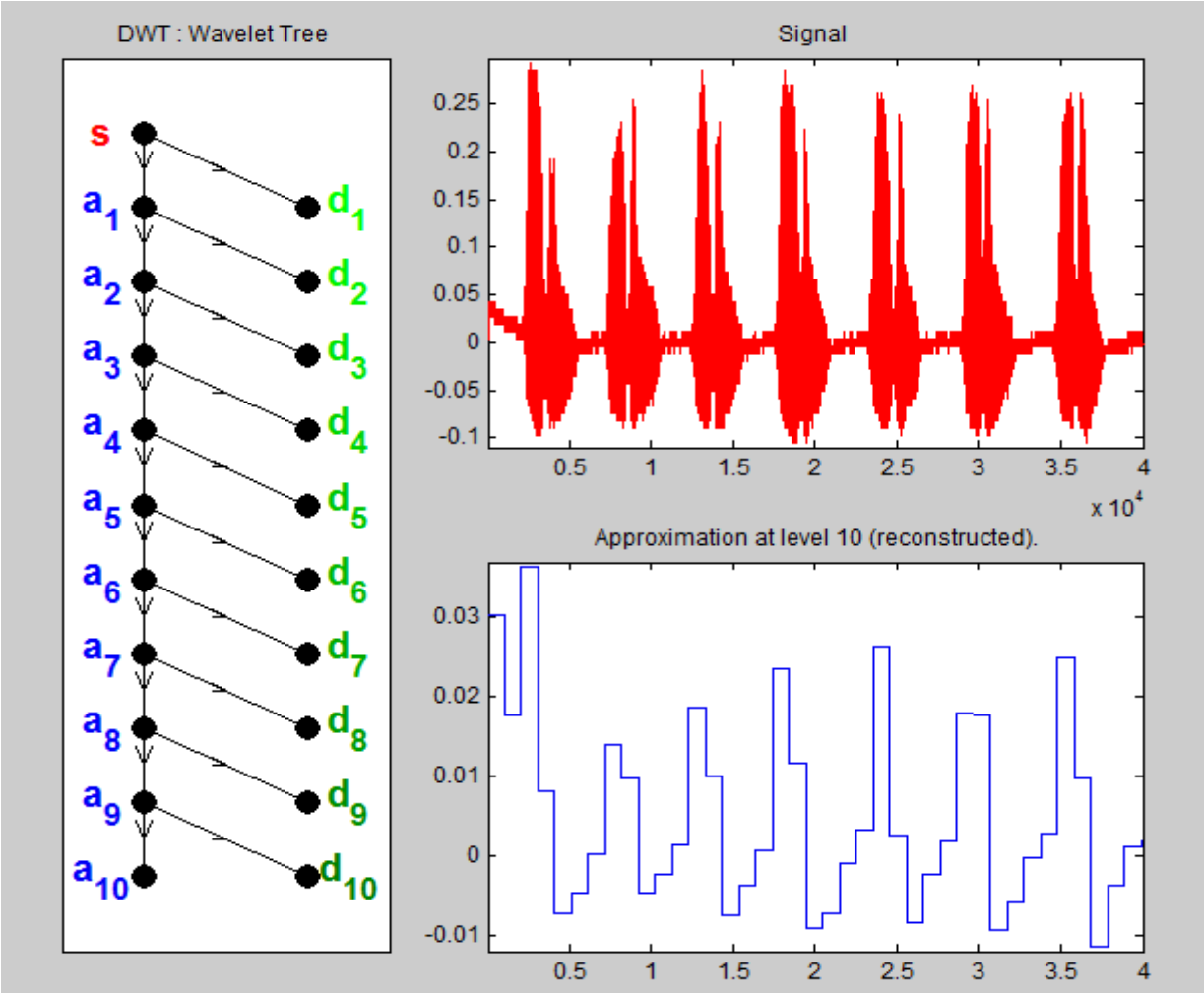


Fig. 8. Original speech signal of Mr."A" with 10 level decomposition and decomposition tree.

L_1 Norm = 1157

L_2 Norm = 10.02

Signal to noise ratio = 5.72 db

1.4.8 DATA CHART FOR MR. "A":

	1st voice	2nd voice	3rd voice
L_1 Norm	491.2	842.8	1157
L_2 Norm	3.51	7.00	10.02
SNR	5.83	5.87	5.72

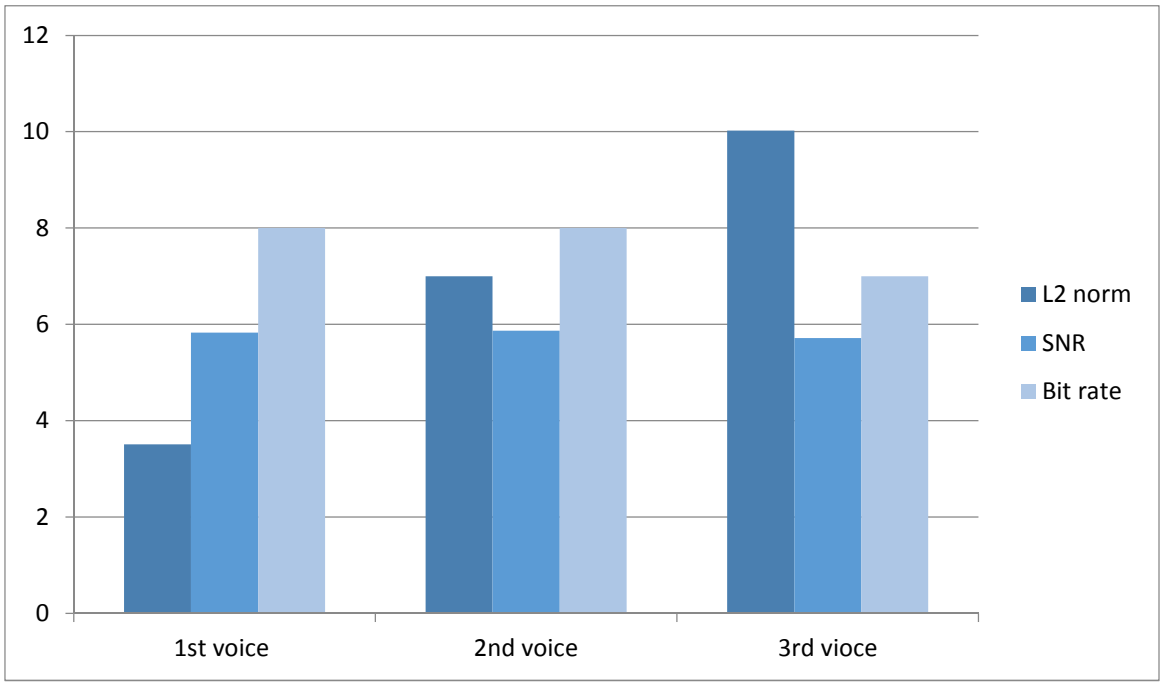


Fig. 9. Three experimental voice data chart of Mr."A".

Summary:

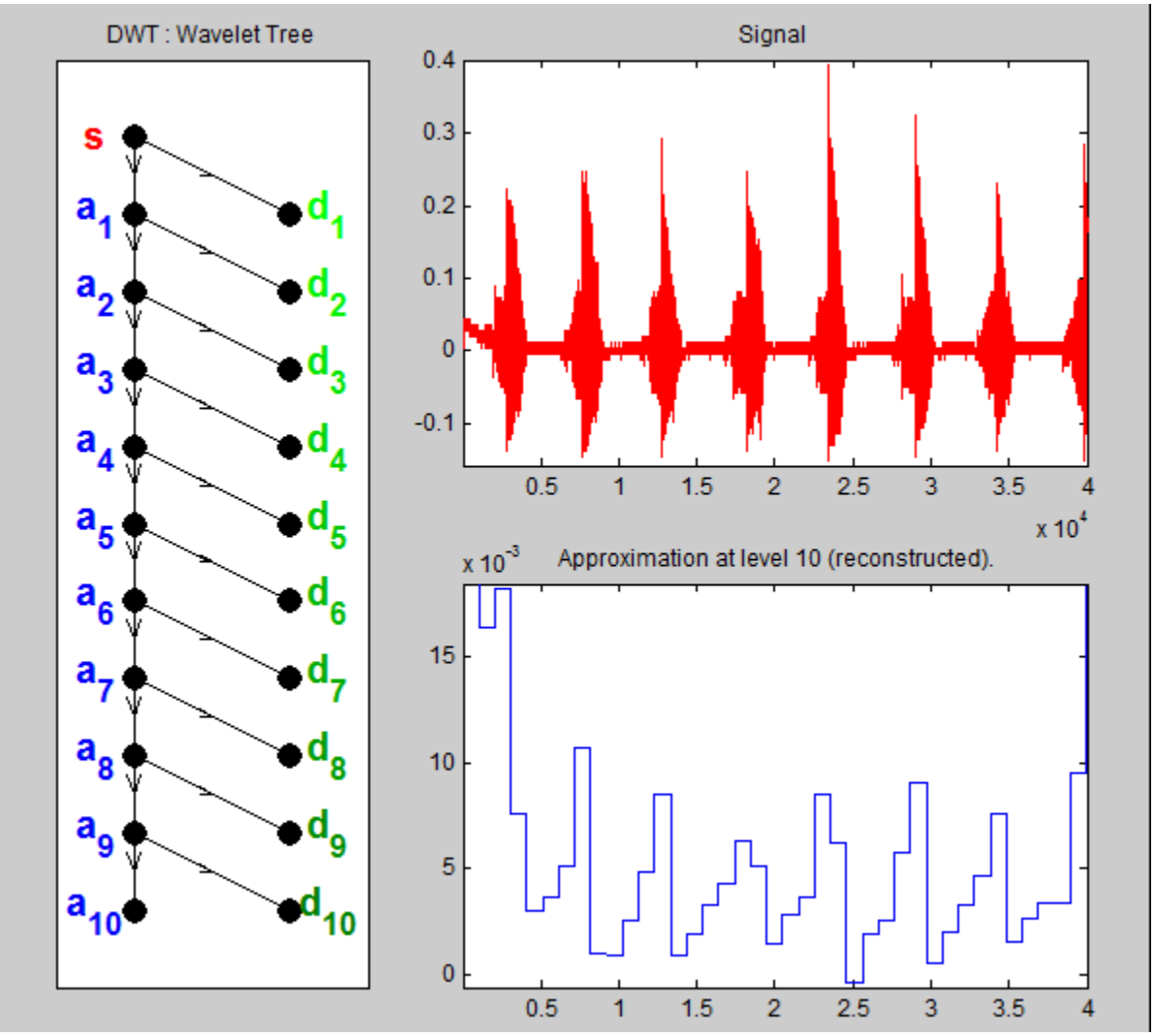
From the above chart it is concluded that the SNR value is minimum for the high bit rate voice(fast voice) than that the slow bit rate voice(slow voice) within the same decibel (db) value. If we increase the volume of vocal chord in different cases then the result will change. With the increase of volume of vocal chord the value of L1 norm and L2 norm is increase respectively. For a same range bit rate voices with different volume, SNR value may be increased or decreased but L1and L2 norm must increased with the increase of volume. In this result we can observed that the 3rd speech has the high volume i.e high db value among the 3 speech and the 1st Speech is faster voice than the other two.

1.4.9 EXPERIMENT 7: ORIGINAL SPEECH SIGNAL OF MR. "B":

CODING OF SPEECH SIGNAL:

Mr. "B" is talking with headphone: Hello, Hello, Hello, Hello, Hello, Hello, Hello, Hello

(8 Times)



X axis→ time (ms), Y axis→

Fig. 10. Original speech signal of Mr ."B" with 10 level decomposition and decomposition tree.

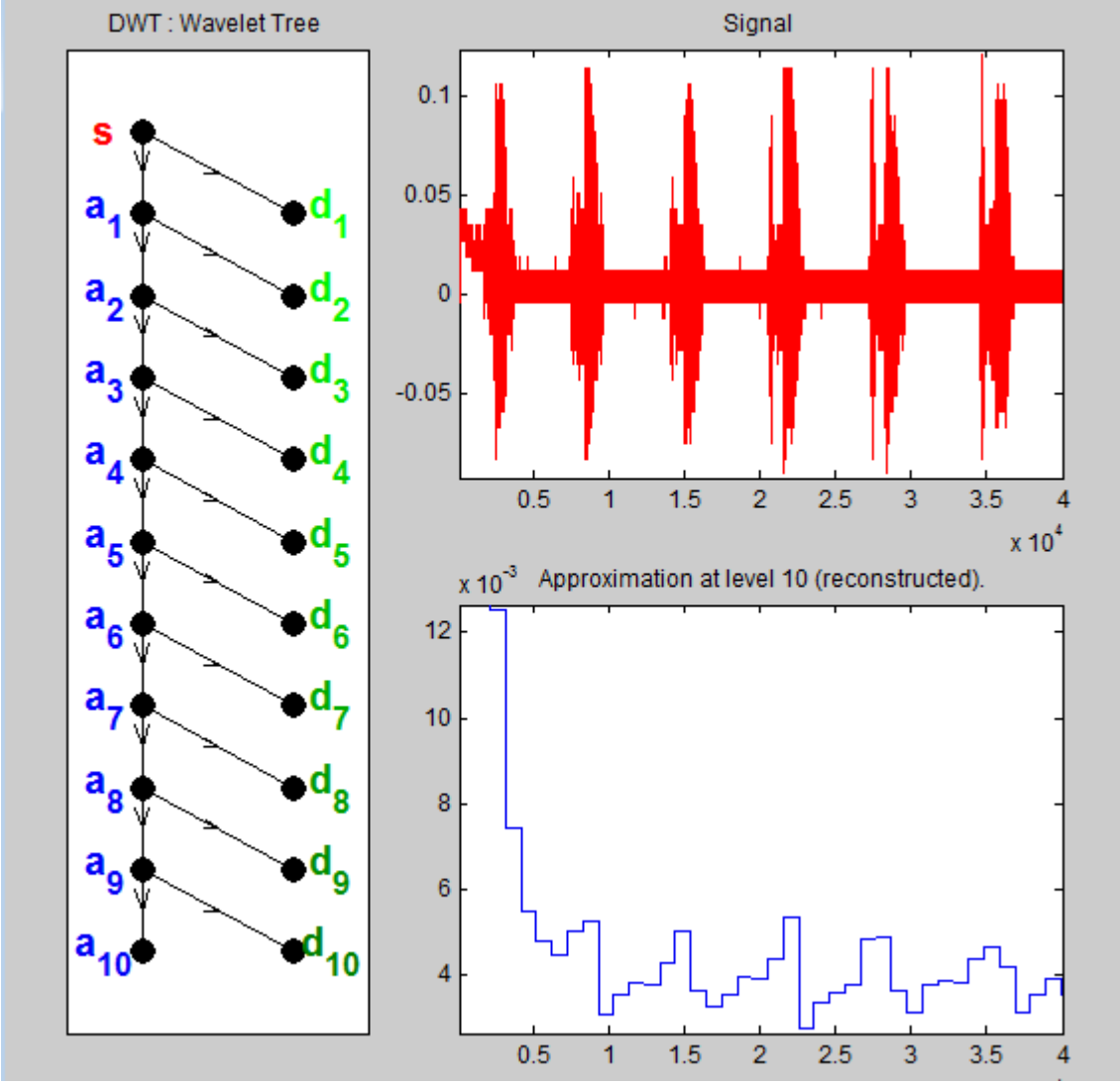
$\mathbf{L_1\ Norm = 789.9}$

$\mathbf{L_2\ Norm = 7.064}$, *Signal to noise ratio = 7.40 db*

1.4.10 EXPERIMENT 8: ORIGINAL SPEECH SIGNAL OF MR. "B":

CODING OF SPEECH SIGNAL:

Mr. "B" is talking with headphone: Hello, Hello, Hello, Hello, Hello, Hello, (6 Times)



X axis→ time (ms), Y axis→

Fig. 11. Original speech signal of Mr. "B" with 10 level decomposition and decomposition tree.

$\mathbf{L_1\ Norm = 471.0}$

$\mathbf{L_2\ Norm = 3.815}$

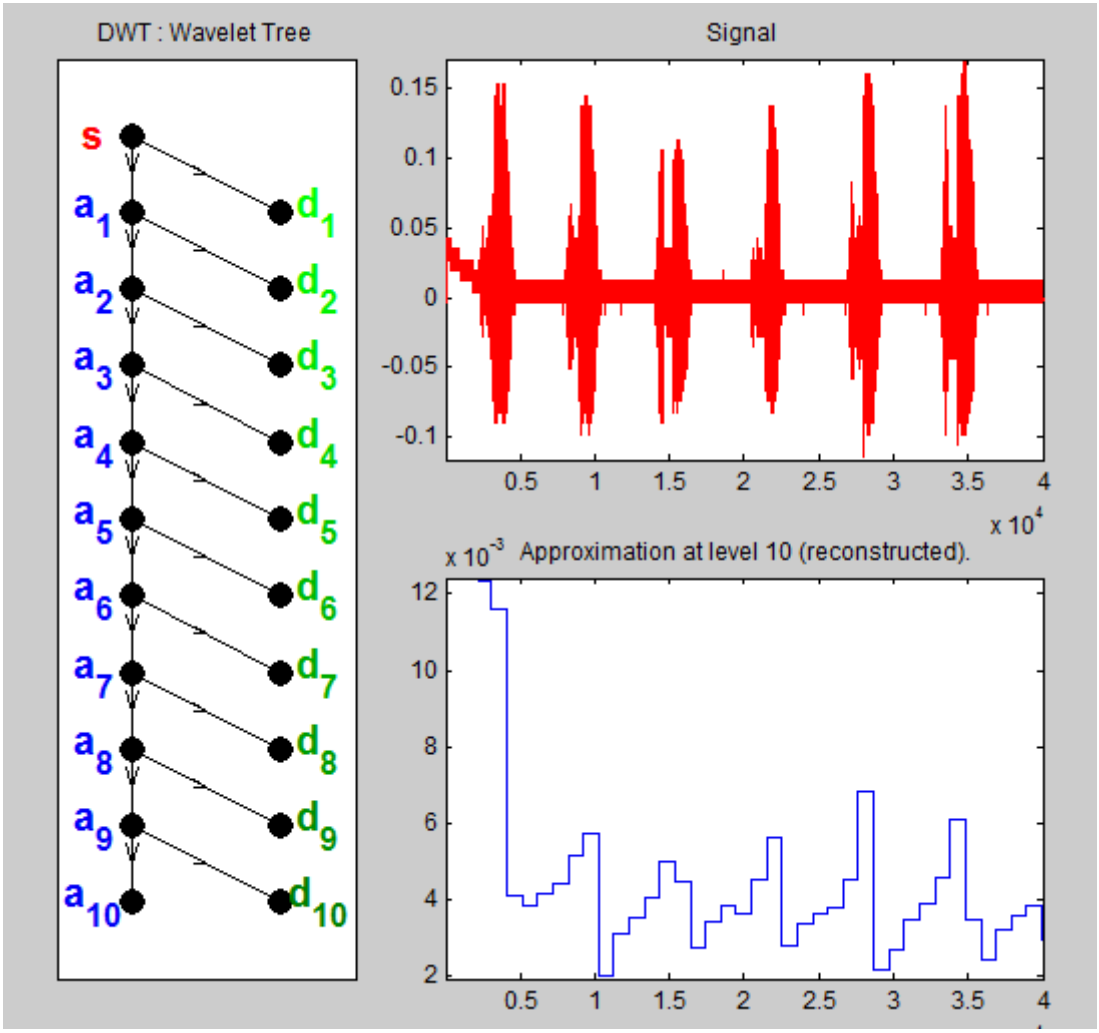
Signal to noise ratio = 7.68 db

1.4.11 EXPERIMENT 9: ORIGINAL SPEECH SIGNAL OF MR. "B":

CODING OF SPEECH SIGNAL:

Mr. "B" is talking with headphone: Hello, Hello, Hello, Hello, Hello, Hello, Hello, and Hello

(6 Times with loud enough)



X axis→ time (ms), Y axis→

Fig. 12. Original speech signal of Mr."B" with 10 level decomposition and decomposition tree.

L_1 Norm = 559.8

L_2 Norm = 4.904

Signal to noise ratio = 7.06 db

1.4.12 DATA CHART FOR MR. "B":

	1st voice	2nd voice	3rd voice
L_1 Norm	789.9	471.0	559.8
L_2 Norm	7.06	3.81	4.90
SNR	7.40	7.68	7.06

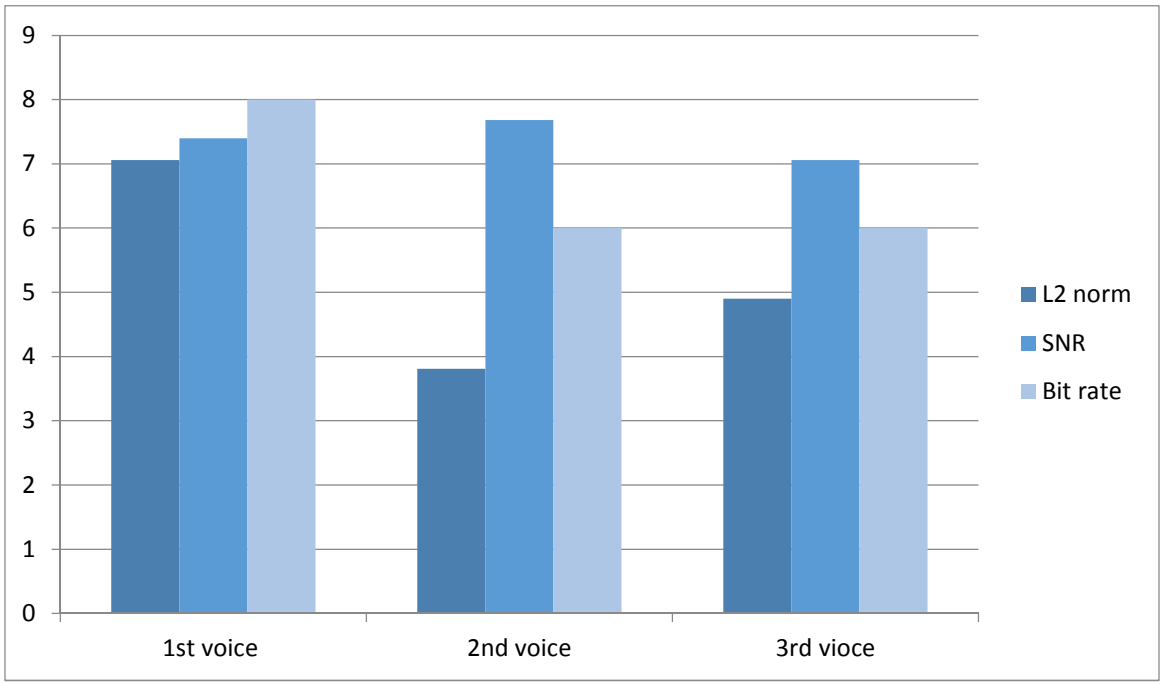


Fig. 13. Three experimental voice data chart of Mr."B".

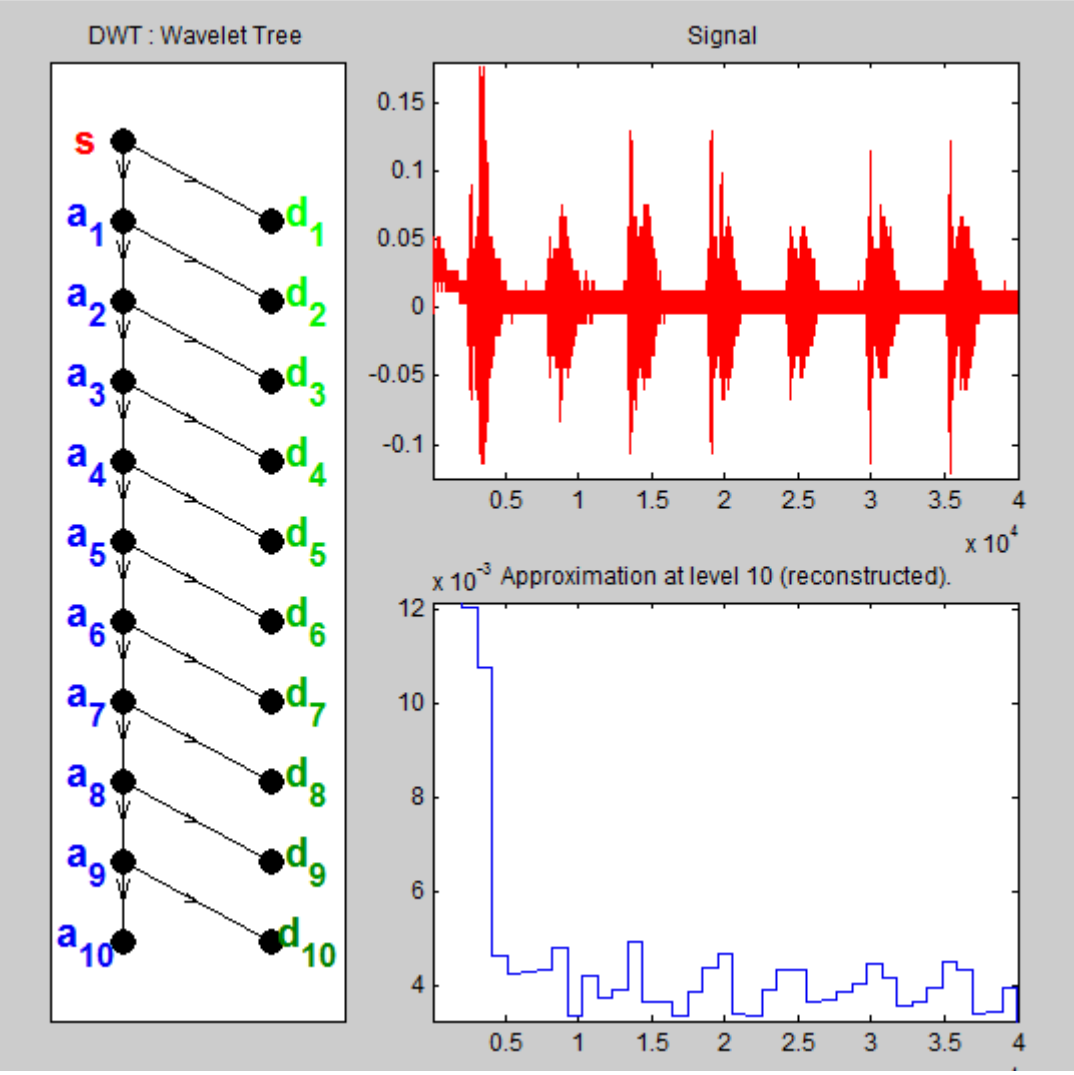
Summary:

From the above chart it is concluded that the SNR value is minimum for the high bit rate voice (fast voice) than that the slow bit rate voice (slow voice) within the same decibel (db) value. If we increase the volume of vocal chord in different cases then the result will change. With the increase of volume of vocal chord the value of L1 norm and L2 norm is increase respectively. For a same range bit rate voices with different volume, SNR value may be increased or decreased but L1and L2 norm must increased with the increase of volume. By this way we can say that the 1st speech has the high volume i.e high db value among the 3 speech and the 3rd Speech is faster voice than the other two.

1.4.13 EXPERIMENT 10: ORIGINAL SPEECH SIGNAL OF MR. "C":

CODING OF SPEECH SIGNAL:

Mr. "C" is talking with headphone: Hello, Hello, Hello, Hello, Hello, Hello, and Hello (7 Times)



X axis→ time (ms), Y axis→

Fig. 14. Original speech signal of Mr."C" with 10 level decomposition and decomposition tree.

L_1 Norm = 463.7

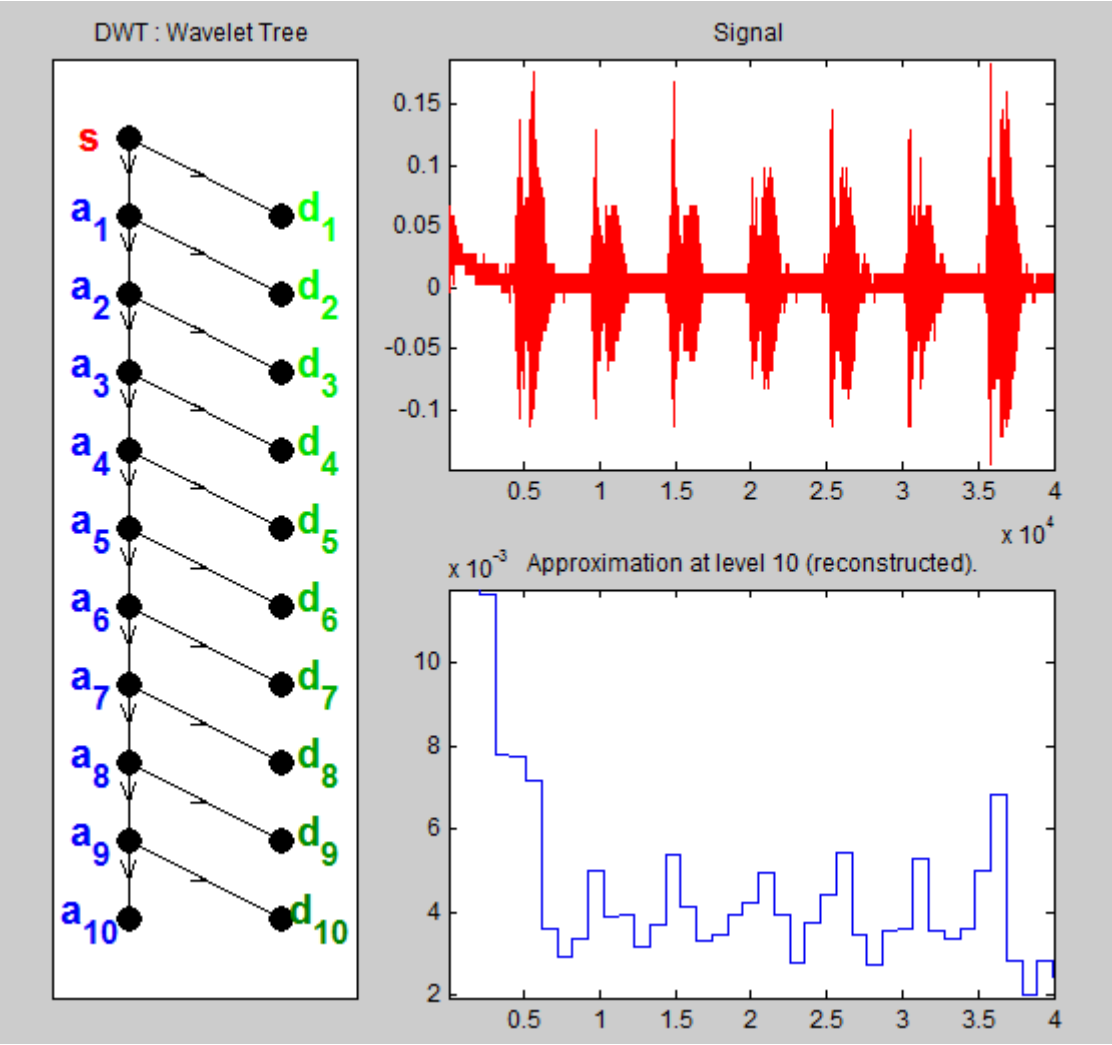
L_2 Norm = 3.546

Signal to noise ratio = 7.367 db

1.4.14 EXPERIMENT 11: ORIGINAL SPEECH SIGNAL OF MR. "C":

CODING OF SPEECH SIGNAL:

Mr. "C" is talking with headphone: Hello, Hello, Hello, Hello, Hello, Hello, and Hello (7 Times)



X axis→ time (ms), Y axis→

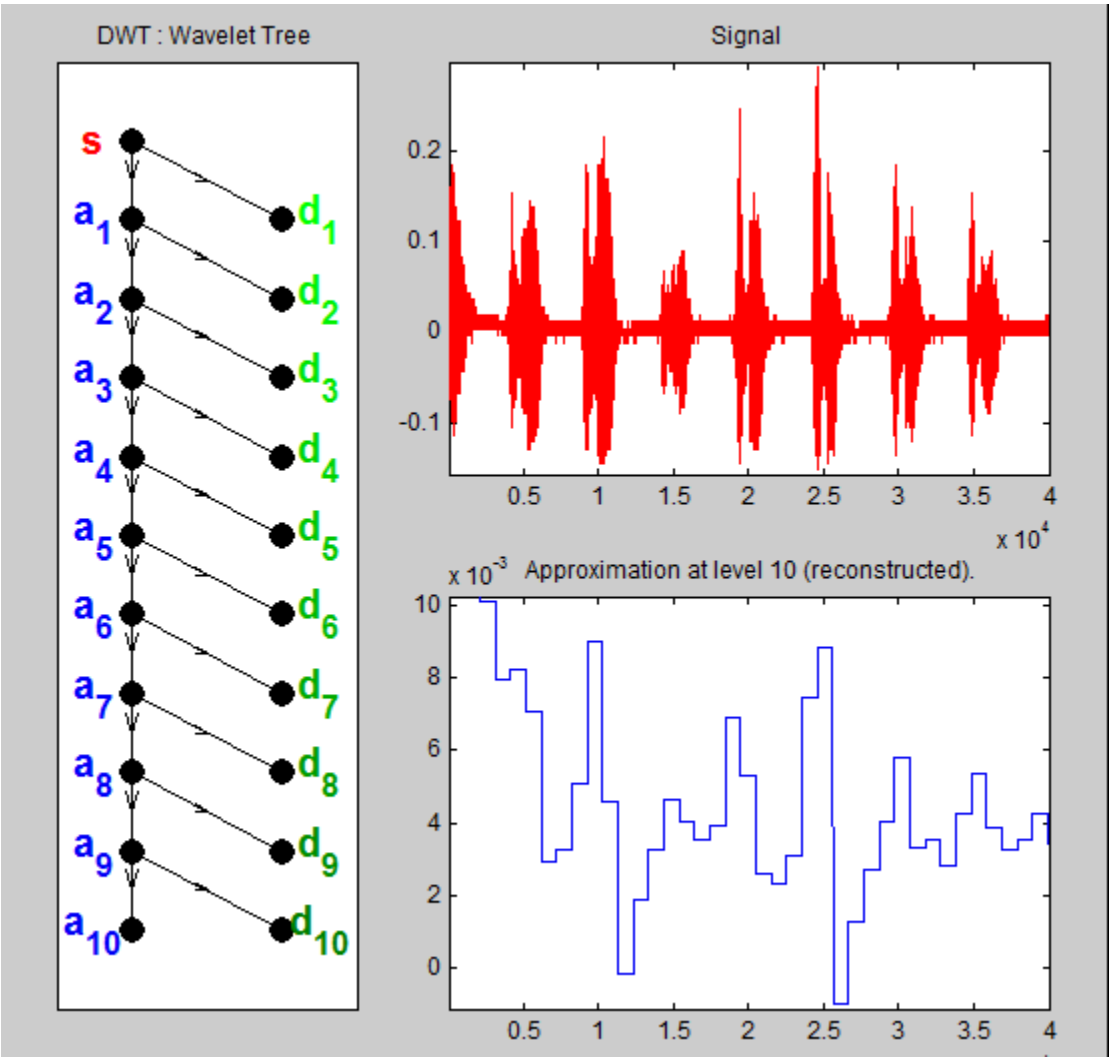
Fig. 15. Original speech signal of Mr. "C" with 10 level decomposition and decomposition tree.
$\mathbf{L_1\ Norm = 561.7}$
$\mathbf{L_2\ Norm = 4.467}$, ***Signal to noise ratio = 6.64 db***

1.4.15 EXPERIMENT 12: ORIGINAL SPEECH SIGNAL OF MR. "C":

CODING OF SPEECH SIGNAL:

Mr. "C" is talking with headphone: Hello, Hello, Hello, Hello, Hello, Hello, Hello, Hello

(8 Times with loud enough)



X axis→ time (ms), Y axis→

Fig. 16. Original speech signal of Mr."C" with 10 level decomposition and decomposition tree.

L_1 **Norm** $= 665.4$

L_2 **Norm** $= 5.774$, ***Signal to noise ratio = 6.00 db***

1.4.16 DATA CHART FOR MR. "C":

	1st voice	2nd voice	3rd voice
L_1 Norm	463.7	561.7	665.4
L_2 Norm	3.54	4.46	5.77
SNR	7.36	6.64	6.00

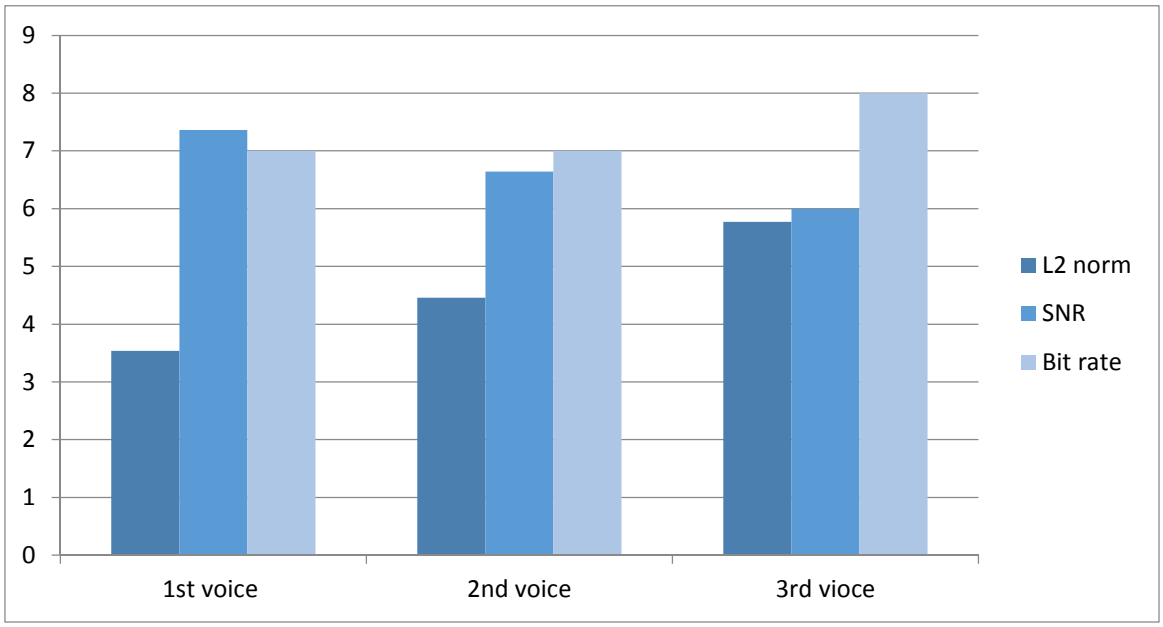


Fig. 17. Three experimental voice data chart of Mr."C".

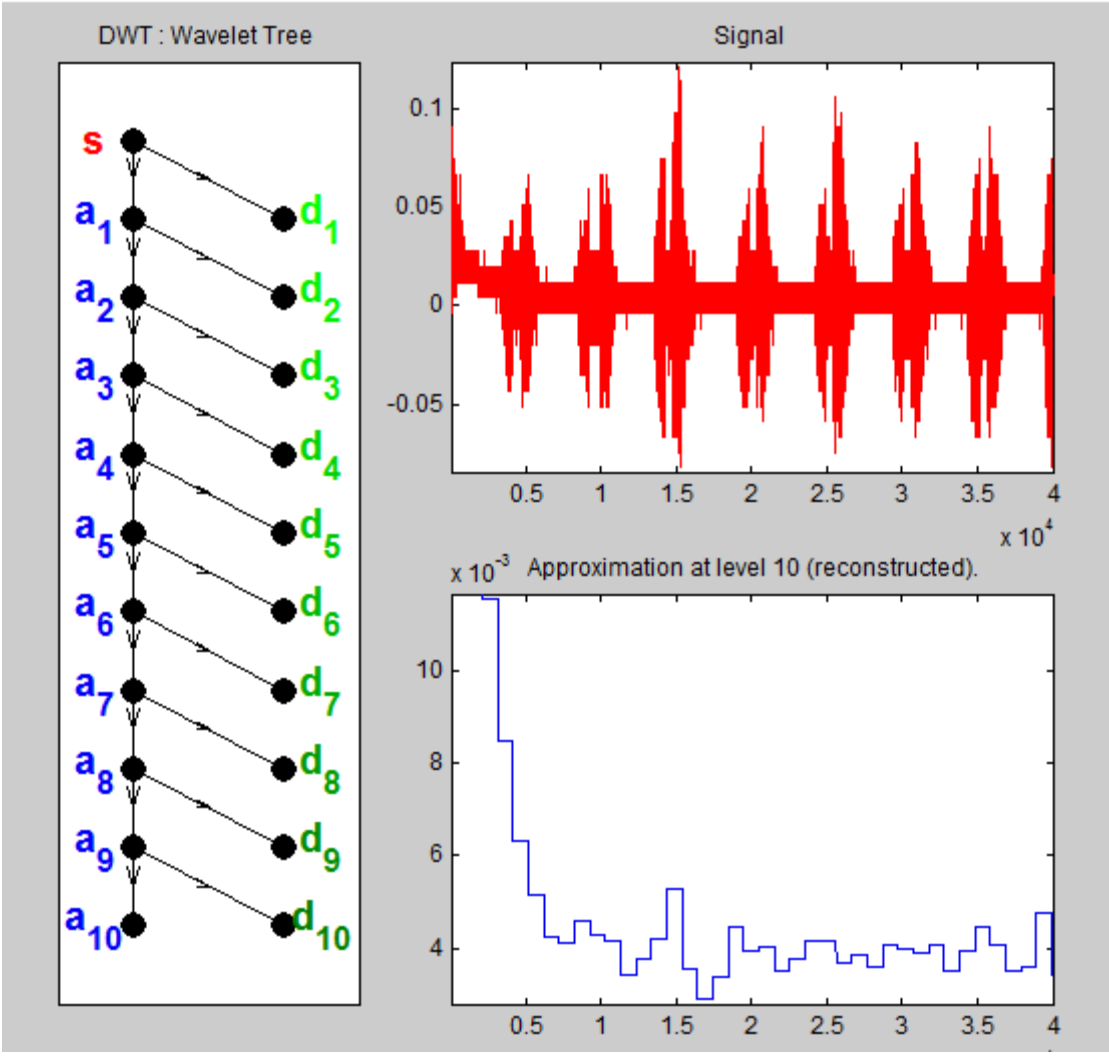
Summary:

From the above chart it is concluded that the SNR value is minimum for the high bit rate voice (fast voice) than that the slow bit rate voice (slow voice) within the same decibel (db) value. If we increase the volume of vocal chord in different cases then the result will change. With the increase of volume of vocal chord the value of L1 norm and L2 norm is increase respectively. For a same range bit rate voices with different volume, SNR value may be increased or decreased but L1and L2 norm must increased with the increase of volume. By this way we can say that the 3rd speech has the high volume i.e high db value among the 3 speech and the 3rd Speech is faster voice than the other two.

1.4.17 EXPERIMENT 13: ORIGINAL SPEECH SIGNAL OF MR. "D":

CODING OF SPEECH SIGNAL:

Mr. "D" is talking with headphone: Hello, Hello, Hello, Hello, Hello, Hello, and Hello (7 Times)



X axis→ time (ms), Y axis→

Fig. 18. Original speech signal of Mr."D" with 10 level decomposition and decomposition tree.

$\mathbf{L_1\ Norm = 450.3}$

$\mathbf{L_2\ Norm = 3.26}$

Signal to noise ratio = 6.906 db

1.4.18 EXPERIMENT 14: ORIGINAL SPEECH SIGNAL OF MR. "D":

CODING OF SPEECH SIGNAL:

Mr. "D" is talking with headphone: Hello, Hello, Hello, Hello, Hello, Hello, and Hello (7 Times with loud enough)

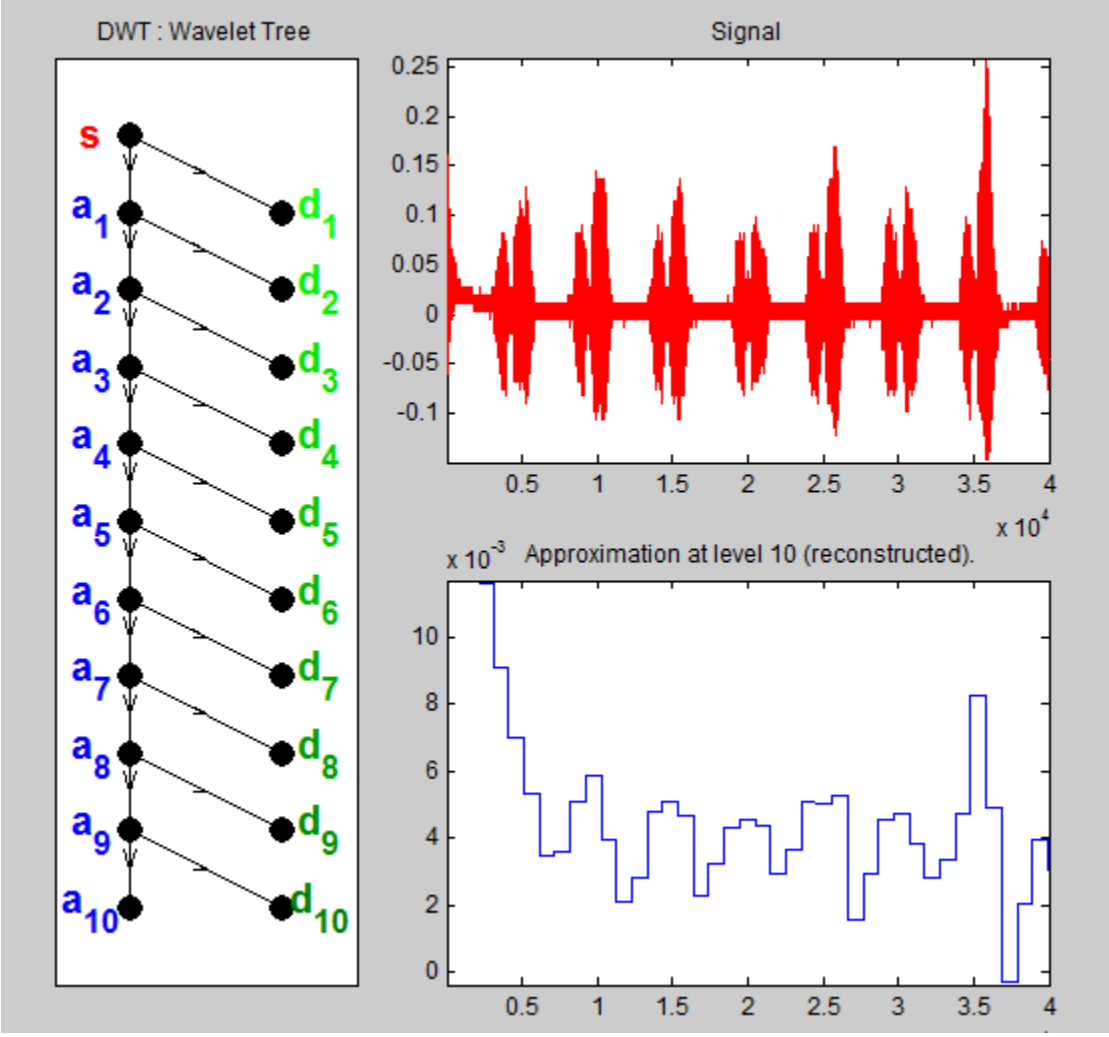


Fig. 19. Original speech signal of Mr."D" with 10 level decomposition and decomposition tree.

$\mathbf{L_1\ Norm = 690.0}$

$\mathbf{L_2\ Norm = 5.708}$

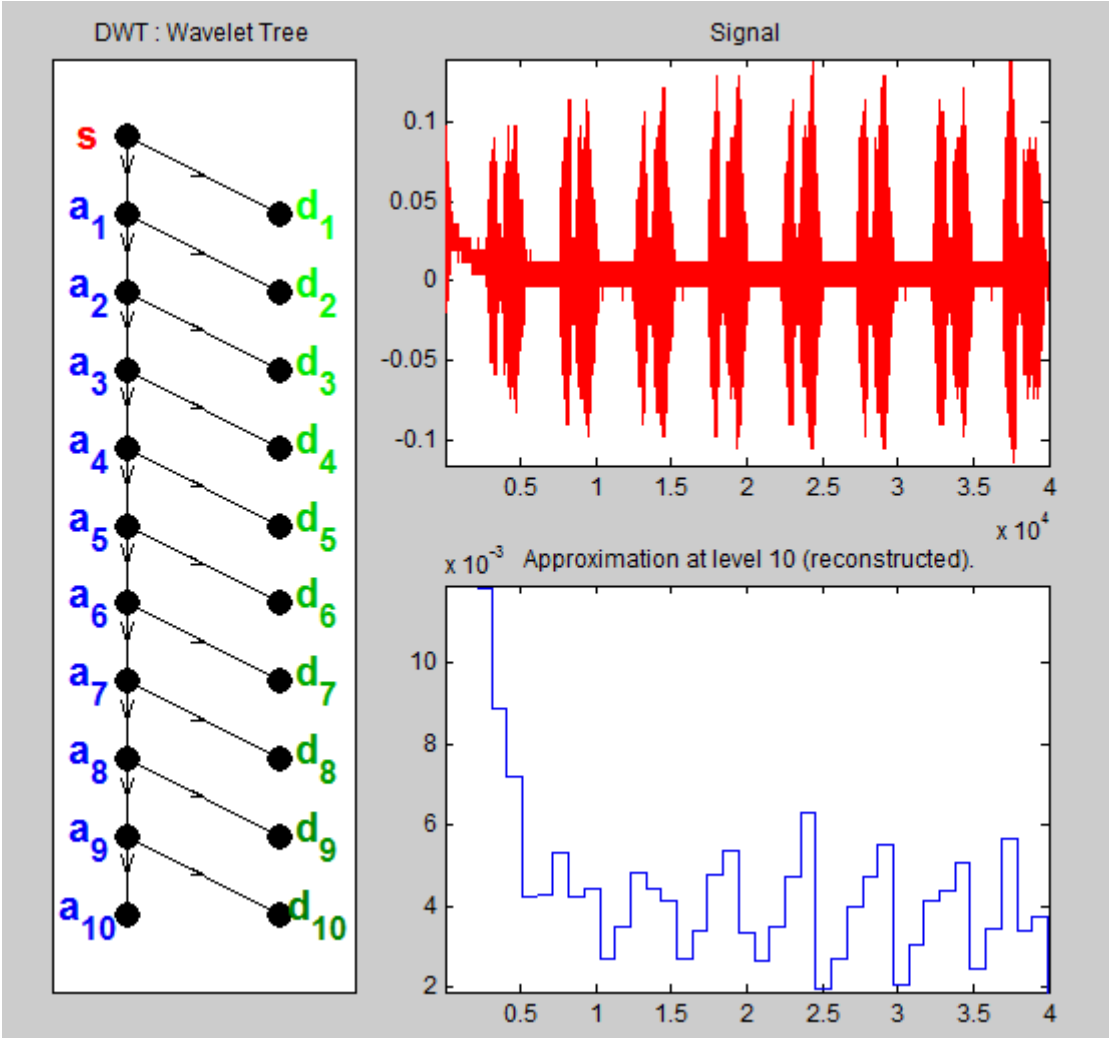
Signal to noise ratio = 6.38 db

1.4.19 EXPERIMENT 15: ORIGINAL SPEECH SIGNAL OF MR. "D":

CODING OF SPEECH SIGNAL:

Mr. "D" is talking with headphone: Hello, Hello, Hello, Hello, Hello, Hello, Hello, and Hello

(8 Times with loud enough)



X axis→ time (ms), Y axis→

Fig. 20. Original speech signal of Mr."D" with 10 level decomposition and decomposition tree.

L_1 **Norm** $= 680.0$

L_2 **Norm** $= 5.23$

Signal to noise ratio = 5.73 db

1.4.20 DATA CHART FOR MR. "D":

	1st voice	2nd voice	3rd voice
L_1 **Norm**	450.3	690.0	680.0
L_2 **Norm**	3.26	5.70	5.23
SNR	6.90	6.38	5.73

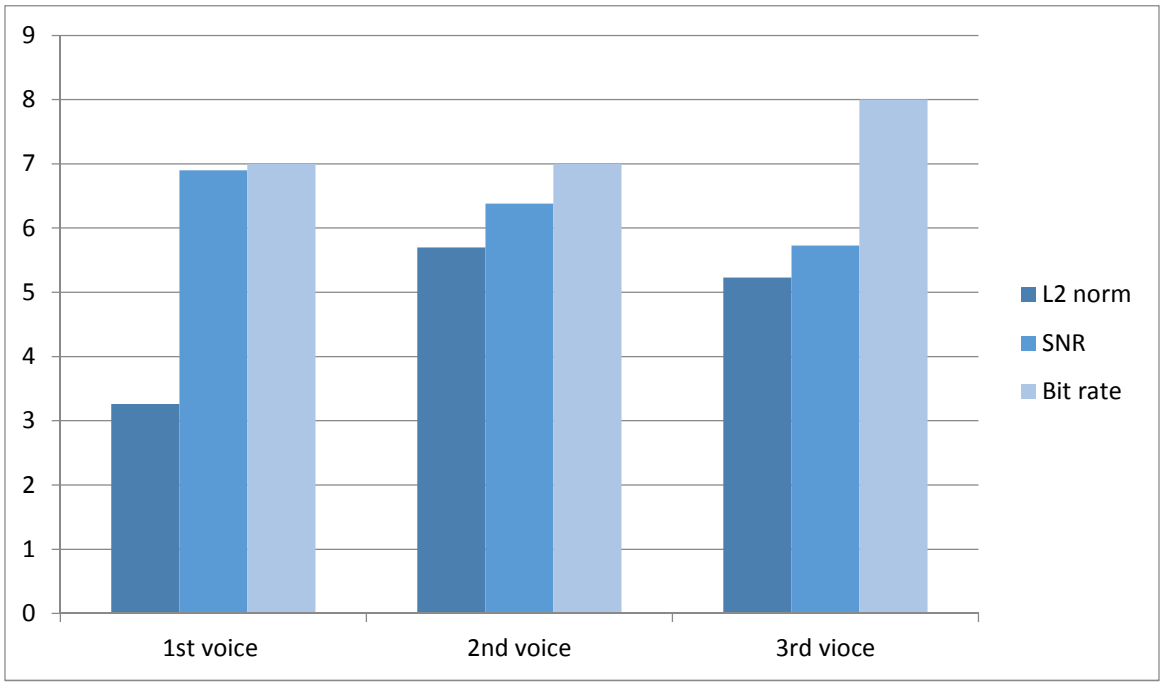


Fig. 21. Three experimental voice data chart of Mr."D".

Summary:

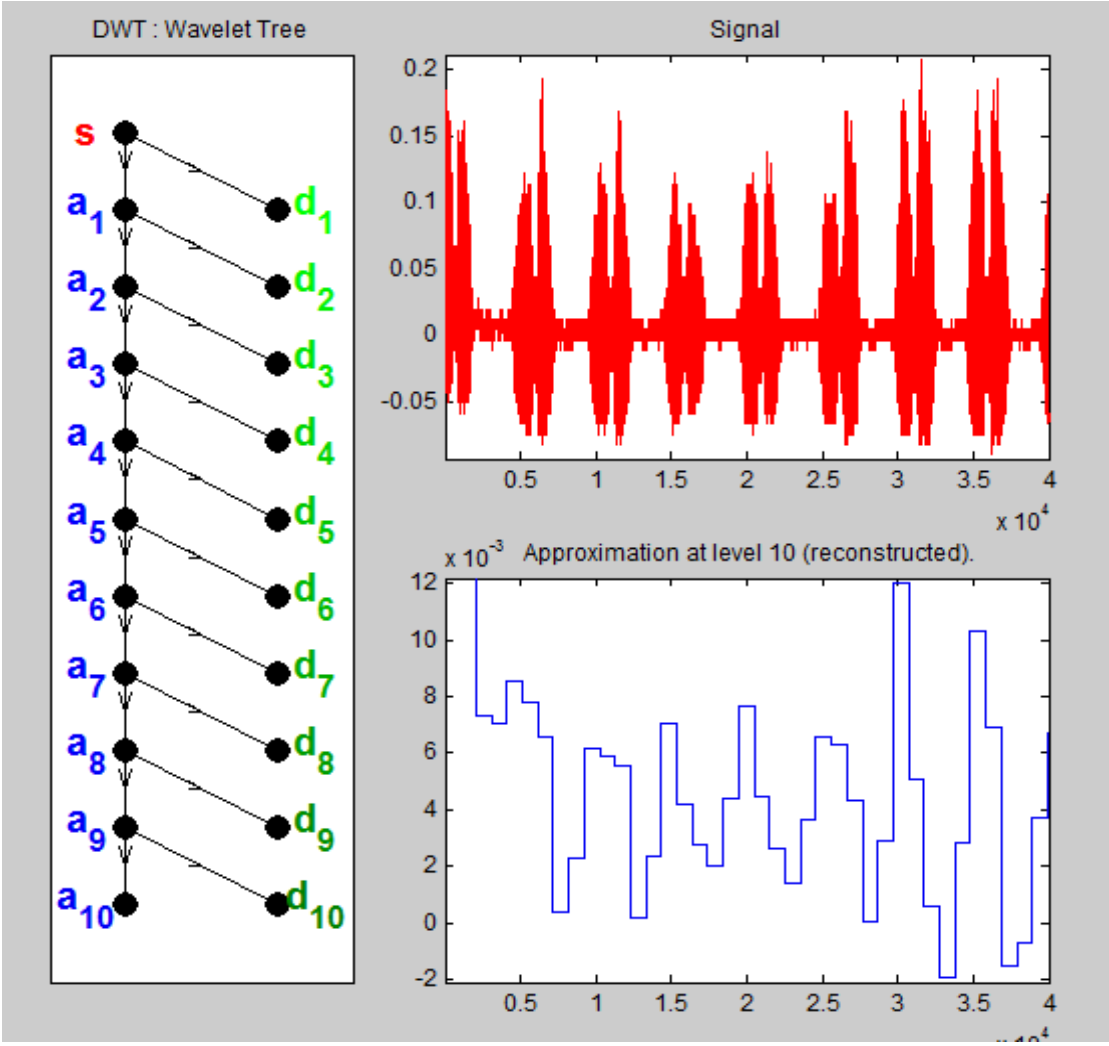
From the above chart it is concluded that the SNR value is minimum for the high bit rate voice (fast voice) than that the slow bit rate voice (slow) within the same decibel (db) value. If we increase the volume of vocal chord in different cases then the result will change. With the increase of volume of vocal chord the value of L1 norm and L2 norm is increase respectively. For a same range bit rate voices with different volume, SNR value may be increased or decreased but L1and L2 norm must increased with the increase of volume. By this way we can say that the 3rd speech has the high volume i.e high db value among the speech and the 3rd Speech is faster voice than the other two.

1.4.21 EXPERIMENT 16: ORIGINAL SPEECH SIGNAL OF MR. "E"

CODING OF SPEECH SIGNAL

Mr. "E" is talking with headphone: Hello, Hello, Hello, Hello, Hello, Hello, Hello, and Hello

(8 Times with loud enough)



X axis→ time (ms), Y axis→

Fig. 22. Original speech signal of Mr."E" with 10 level decomposition and decomposition tree.

$\mathbf{L_1\ Norm = 735.6}$

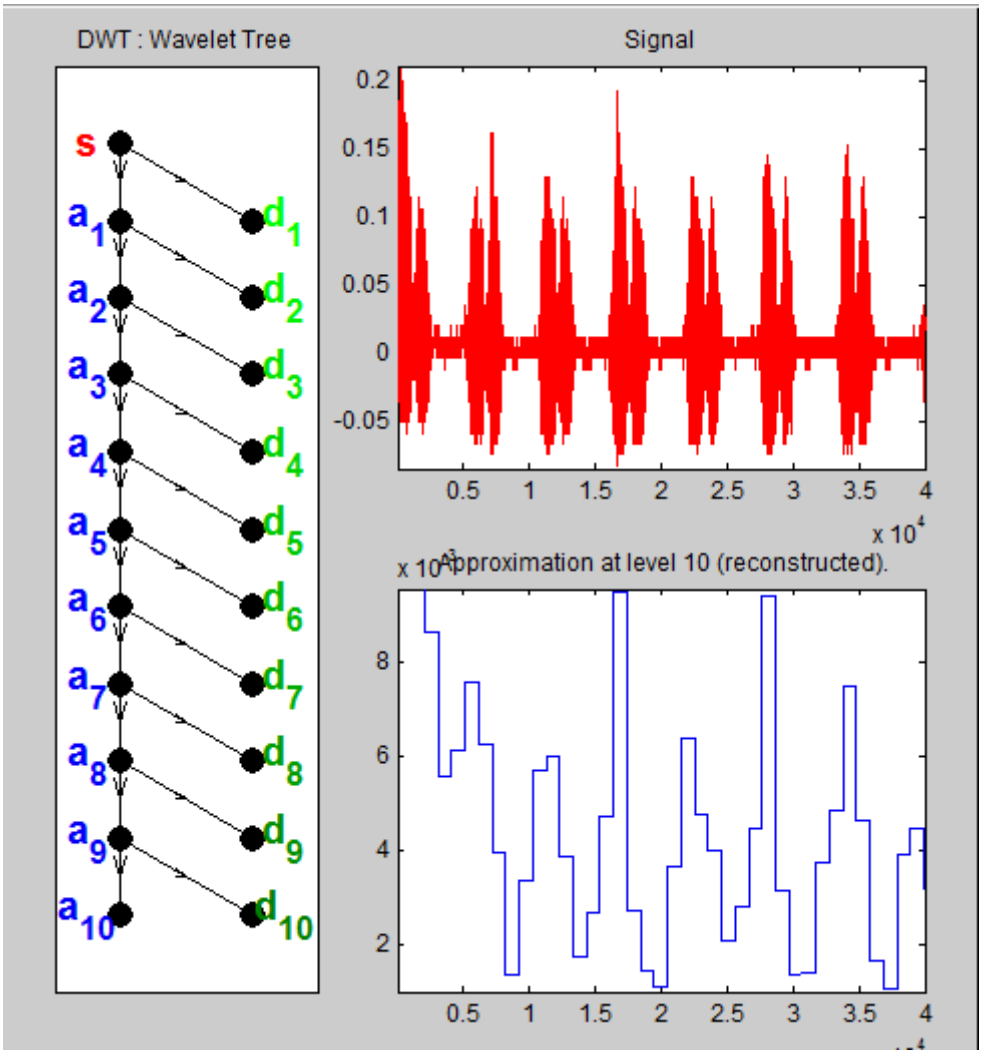
$\mathbf{L_2\ Norm = 5.91}$

Signal to noise ratio = 2.33 db

1.4.22 EXPERIMENT 17: ORIGINAL SPEECH SIGNAL OF MR. "E"

CODING OF SPEECH SIGNAL:

Mr. "E" is talking with headphone: Hello, Hello, Hello, Hello, Hello, Hello, Hello, and Hello (7 Times)



X axis→ time (ms), Y axis→

Fig. 23. Original speech signal of Mr."E" with 10 level decomposition and decomposition tree.

$\mathbf{L_1\ Norm = 603.3}$

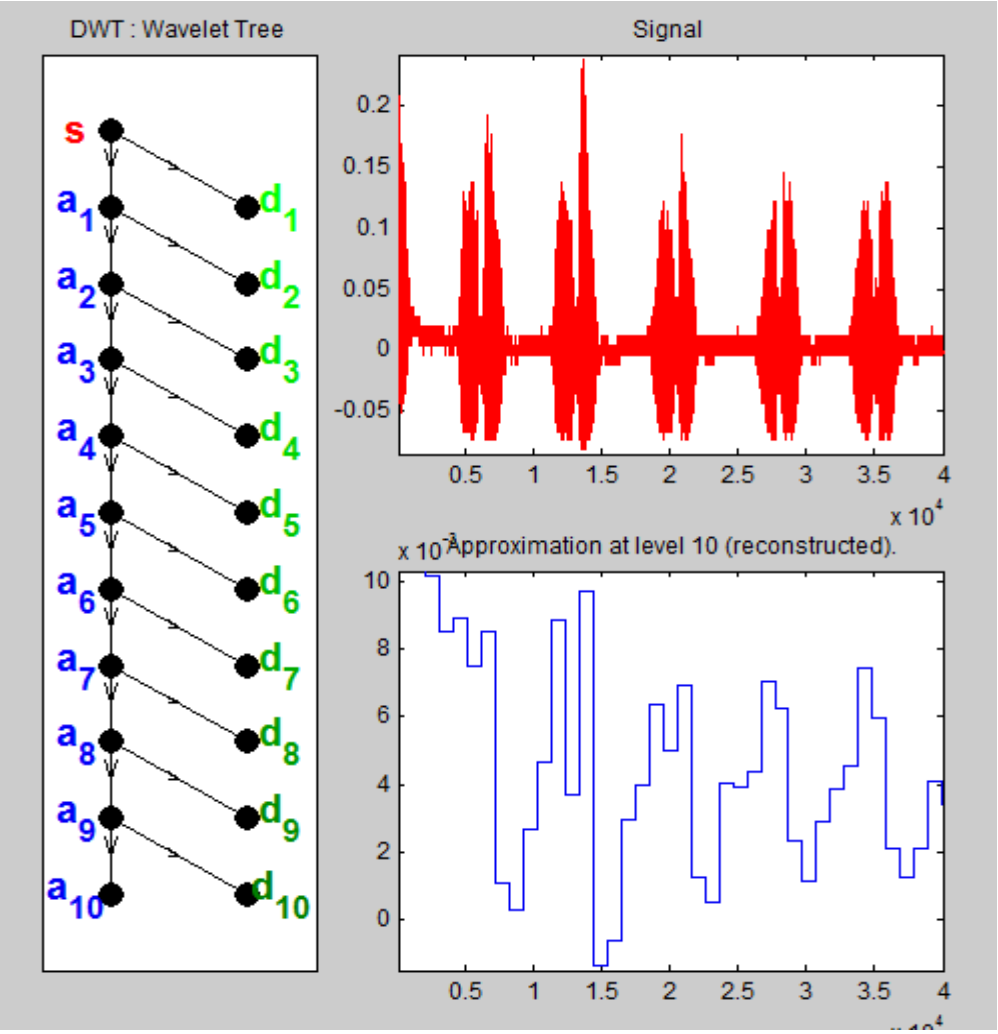
$\mathbf{L_2\ Norm = 4.991}$

Signal to noise ratio = 2.65 db

1.4.23 EXPERIMENT 18: ORIGINAL SPEECH SIGNAL OF MR. "E"

CODING OF SPEECH SIGNAL:

Mr. "E" is talking with headphone: Hello, Hello, Hello, Hello, Hello, Hello, Hello, and Hello (5.5 Times)



X axis→ time (ms), Y axis→

Fig. 24. Original speech signal of Mr."E" with 10 level decomposition and decomposition tree.

L_1 Norm = 601.5

L_2 Norm = 5.024

Signal to noise ratio = 2.86 db

1.4.24 DATA CHART FOR MR. "E":

	1st voice	2nd voice	3rd voice
L_1 Norm	735.6	603.3	601.5
L_2 Norm	5.91	4.99	5.02
SNR	2.33	2.65	2.86

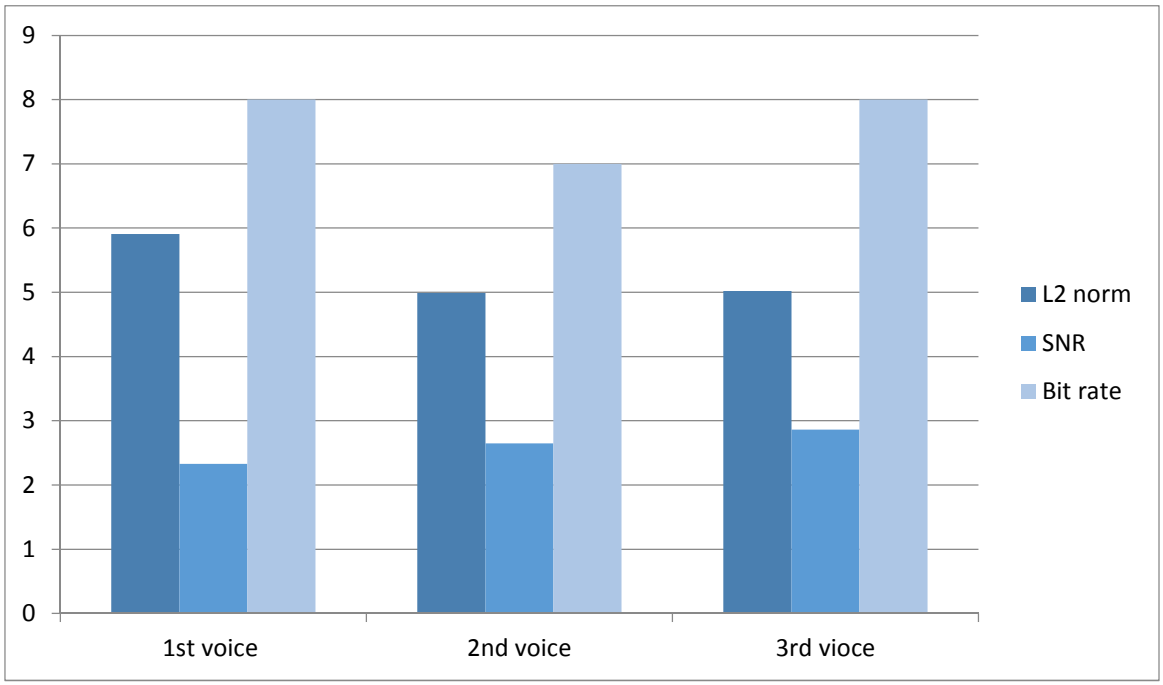
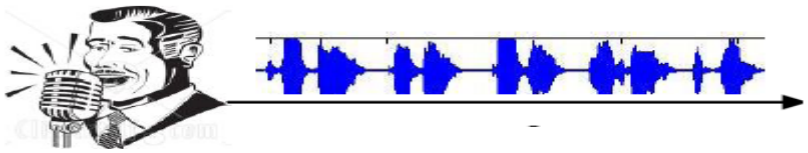


Fig. 25. Three experimental voice data chart of Mr."E"

Summary:

From the above chart it is concluded that the SNR value is minimum for the high bit rate voice (fast voice) than that the slow bit rate voice (slow voice) within the same decibel (db) value. If we increase the volume of vocal chord in different cases then the result will change. With the increase of volume of vocal chord the value of L1 norm and L2 norm is increase respectively. For a same range bit rate voices with different volume, SNR value may be increased or decreased but L1and L2 norm must increased with the increase of volume. By this way we can say that the 1st speech has the high volume i.e high db value among the 3 speech and the 1st Speech is faster voice than the other two.

2 RESULTS AND DISCUSSIONS



All discursion for continuous English speaking

TABLE:01

Person→ Cal↓	Voice of MR."A"			Voice of MR."B"			Voice of MR."C"			Voice of MR."D"		
	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd
L_1 Norm	481.8	587.5	570.1	789.9	471.0	559.8	463.7	561.7	665.4	450.3	690	680
L_2 Norm	3.84	5.05	4.78	7.06	3.81	4.90	3.54	4.46	5.77	3.26	5.70	5.23
SNR	6.81	6.00	5.81	7.40	7.68	7.06	7.36	6.64	6.00	6.90	6.38	5.73

TABLE: 02

Person→ Cal↓	Voice of Mr. "E"		
	1st	2nd	3rd
L_1 Norm	735.6	603.3	601.5
L_2 Norm	5.91	4.99	5.024
SNR	2.33	2.65	2.86

Fig. 26. Data chart for all of speech analysis.

2.1 RESULT

As shown in table 6 a speech files spoken in English language is recorded for only male. The effects of varying threshold value on the speech signals in terms of SNR and compression score were observed for different cases. There are many factors which affects the wavelet based speech Coder's performance, mainly what compression ratio could be achieved at suitable SNR value. To improve the compression ratio of wavelet-based coder, we have to consider that it is highly speaker dependent and varies with his age and gender. That is low speaking speed because high compression ratio with high value of SNR and high speaking speed cause low compression ratio with low value of SNR. Also the detection of volume depends on the value of $\mathbf{L_1}$ and $\mathbf{L_2}$ Norm. The High volume speaking cause high $\mathbf{L_1}$ and $\mathbf{L_2}$ Norm on the other hand low volume speaking cause low $\mathbf{L_1}$ and $\mathbf{L_2}$ Norm respectively. All of signals are analyzed in Haar Wavelet.

3 CONCLUSION

We are tried to observe voice activity or detect the voice activity by calculating and comparing the mathematical term$\mathbf{L_1}$, $\mathbf{L_2}$ Norm. In this study we also calculated SNR for each signal. In this study we have analyze different speech signal by Haar wavelet with 10 decomposition level. Performance of the wavelet coder is tested on male speech signals of duration 5 Sec. Results illustrate that with the help of wavelet we can analysis and detects the voice activity. For getting accurate more result it is require to high machineries and farther study.

REFERENCES

[1] Furui S., (1985), "Digital Speech Processing", Tokai University Pub, (in Japan).

[2] Abdallah I.,et al., (1997), "Robust Speech/Non Speech Detection in Adverse Condition Using an Entropy Based Estimator", *13th International Conference on Digital Signal Processing*., Vol. 2, No. 3, pp. 757-760.

[3] Kadambe S., Boudreaux-Bartels G.F., (1992), "Application of the Wavelet Transform for Pitch Detection ofSpeech Signals", *Institute of Electrical and Electronics Engineers Communications Information. Information Theory.*, Vol.38, no.2, pp.917-924.

[4] Chen J-F., Ser W., (2000), "Speech Detection Using Mircophone Array", *Electronic Letters*., Vol.36, no. 2, pp 181-182.

[5] Freeman D. K., Southcott C. B., Boyd I., and Cosier G., (1989), "A voice activity detector for pan-European digital cellular mobile telephone service", in *Proceedings Institute of Electrical and Electronics Engineers Trans. Audio, Speech and Language Processing*, Vol. 19, No. 3, pp 600-613.

[6] Sangwan A., Chiranth M. C., Jamadagni H. S., Sah R., Prasad R. V., and Gaurav V., (2002), "VAD techniques for real-time speech transmission on the Internet", *Proceedings Institute of Electrical and Electronics Engineers Transactions on Communications*, Vol. 52, No. 12, *pp*. 2154- 2164.

[7] Itoh K., and Mizushima M., (1997), "Environmental noise reduction based on speech/non-speech identification for hearing aids," *in Proceedings International conference on acoustic, speech and signal processing*., Vol. 23, No. 1, pp. 419–422.

[8] Vlaj D., Kotnik B., Horvat B., and Kacic Z., (2005), "A computationally efficient mel-filter bank VAD algorithm for distributed speech recognition systems," *The European Association for Signal Processing J. Appl. Signal Process*., Vol. 18, *No.* 3, *pp. 487-497.*

[9] Enqing D., Heming Z., and Yongli L.,(2002), "Lowbit and variable rate speech coding using local cosine transform", *in Proceedings International conference on acoustic, speech and signal processing* ., Vol.1, No. 1, pp. 423–426.

[10] Kaneda Y., (1990), "Speech Period Detection Using a Microphone Array under Noisy Environments", *Trans. Institute of Electrical and Electronics Engineers, Vol.* 73, No. 8, pp 1391-1398.

[11] Agbinya J I., (1996) "Discrete Wavelet Transform Techniques in Speech Proccesing." *international technical conference of Institute of Electrical and Electronics Engineers.*, Vol.2, No. 8, pp. 514-519.

[12] Kiyohara K., Kaneda Y., et al., (1997), "A Mircophone Array System for Speech Recognition", *In Proceedings Institute of Electrical and Electronics Engineers Int'l Conference on Acoustics, Speech, & Signal Processing*, Vol.1, No. 2, pp. 215-218.