Database Development and Comparative Analysis of Isolated, Spoken Hindi Hybrid Paired Words & Spoken Hindi Partially Reduplicated Words Using Endpoint Detection

VARSHA GUPTA[1]
Department of ECE, Dehradun Institute of Technology, Mussoorie Diversion Road
Dehradun,Uttarakhand-248009, India
E-mail:varshag07@gmail.com
www.dit.edu.in

ANUJ KUMAR SHARMA[2]
Assistant Professor
Department of ECE, Dehradun Institute of Technology, Mussoorie Diversion Road
Dehradun, Uttarakhand-248009, India
E-mail: shaanuj@gmail.com
www.dit.edu.in

ANAND SINGH[3]
Assistant Professor
Department of ECE, Dehradun Institute of Technology, Mussoorie Diversion Road
Dehradun, Uttarakhand-248009, India
E-mail:singh01anand01@gmail.com
www.dit.edu.in

Abstract

This paper describes the database development of three different types of spoken words Isolated, Spoken Hindi Hybrid Paired Words and Spoken Hindi Partially Reduplicated words. The main focus is on the comparative analysis of these Words by using endpoint detection algorithm. The development is done in the robust environment. Different male and female speakers are selected and recording is done in normal mode. Endpoint detection algorithm is used for the analysis purpose of various parameters such as Total duration of time, Number of samples, Root mean square value and Mean power (Intensity) in air of speech signal.

Keywords: Isolated Words, Spoken Hindi Hybrid Paired Words (SHHPW), Spoken Hindi Partially Reduplicated Words (SHPRW), Number of Samples, Duration, Endpoint Detection, Mean Power (intensity) in air, Root mean square value (RMS).

1. Introduction

Speech recognition by machine is one of the most attractive areas for research from last many decades. Basically speech recognition is process of automatic extracting and determining linguistic information conveyed by a speech wave using computers [1]. Basically it is the way of converting the spoken word into the text. This system takes an utterance of speech signal as input and converts it into a text sequence similar to information being conveyed by the input data. Pre-Processing (Front-End) of Speech Signal is very essential in the applications where silence or background noise is completely undesirable [12]. Applications like Speech and Speaker Recognition needs efficient feature extraction techniques from speech signal where most of the voiced part contains Speech or Speaker specific attributes. Endpoint Detection [13, 14] as well as silence removal is well known techniques adopted for many years for this and also for dimensionality reduction in speech that facilitate the system to be computationally more efficient. This type of classification of speech into voiced or silence/unvoiced [16] sounds finds other applications mainly in Fundamental Frequency Estimation, Formant Extraction or Syllable

Marking, Stop Consonant Identification and End Point Detection. Speech signal is one the most complex signal to deal with. In addition to the inherent physiological Complexity of the human production system differs from one speaker to another. If we talk about the spoken words, then it is of different types of short words, moderate words and long words. But spoken words have an edge over short and long words in terms of misrecognition rate, pre-processing time, computation time and requirement of large memory space for storing speech templates. Paired words have gap between two words, which acts as a speech code and can play significant role in recognition process [9].

Now if we talk about the isolated words then in that type of words, no gap is available which is a major parameter of enhancement of security of speech signal, while in Hindi hybrid paired words example- Kathin-Imtihaan we have gap between these two words they are called Spoken Hindi Hybrid Paired Words (SHHPW) in which we uses words from different languages but one word is always from Hindi language and the other one may be from different languages e.g. English, Urdu & Farsi etc. While in Hindi partially reduplicated words example – Kaam-Vaam, again we have gap between these two words and in these type of words only first letter of the words are different and rest are same.

Here we are doing comparative analysis of different type of words spoken by male and female both. Endpoint detection is used here for the analysis purpose. This technique is also called Voice Activity Detection.

2. Database Development and Proposed Algorithm

Ten words from different categories are used for the database purpose. Some of them are Bharat, Sunday, Ek, Hindi etc. from Isolated category and Kathin-Imtihaan, Khas-Aadmi, Achi- Gazal, Nayi-Mohabbat etc. from SHHPW category and Kaam-Vaam, Chai-Shai, Naam-Vaam, Aasan- Vaasan etc. from SHPRW. Ten male and ten female speakers participated in recording session. Normal mode is taken for recording session. Stereo headset H 250 with frequency response 20Hz-20 KHz, input impedance of headset is 32 ohm while input sensitivity of microphone is 2.2kohm.11025 Hz sampling frequency with 16bit PCM, mono selected for storing .wav files direct to the computer. Database contains 30 words, simulated in 1 emotion and 20 speakers. This database is enough for comparative analysis purpose between these types of words.

Figure-1 shows a block diagram of a endpoint detection process. Input speech signals are partitioned into frames using window function and parameters such as energy and zero-crossing rates are extracted. During the non speech period of several milli-seconds, a few background thresholds are evaluated. Which are adjusted continuously using the following non speech frames. Each frame is categorized into speech and non-speech using the background thresholds. Endpoint detection logic extracts the start and end point information using the speech/nonspeech class information of each frame, the number of successive frames of each class, and the parameter values. The aim of speech detection is to separate acoustic events of interest (e.g., speech to be processed) in a continuously recorded signal from other parts of the signal (e.g., background). The need for speech detection occurs in many applications in telecommunications. For example, in analog multichannel transmission systems, a technique called time-assignment speech interpolation (TASI) is often used to take advantage of the channel idle time by detecting the presence of a talker's speech and assigning an unused channel only when speech is detected in order to allow more customer services than the channels would generally provide [10]. With TASI, a transmission medium with 96 voice channels can serve 235 customers; a factor of 2.5 more is served because of long pauses in fluent speech. For automatic speech recognition, Endpoint detection is required to isolate the speech of interest so as to be able to create a speech pattern template [11].

3. Experiment and Results

MATLAB 7.8.0 and PRATT Version 5.3 software platform is used to perform the experiment. Figure 2, 3, 4(a) & Figure 2, 3, 4(b) shows the normal speech waveforms and end point detected waveforms of isolated words, Spoken Hindi Hybrid Paired Words and Spoken Hindi Partially Reduplicated Words of male and female speakers in normal mode.

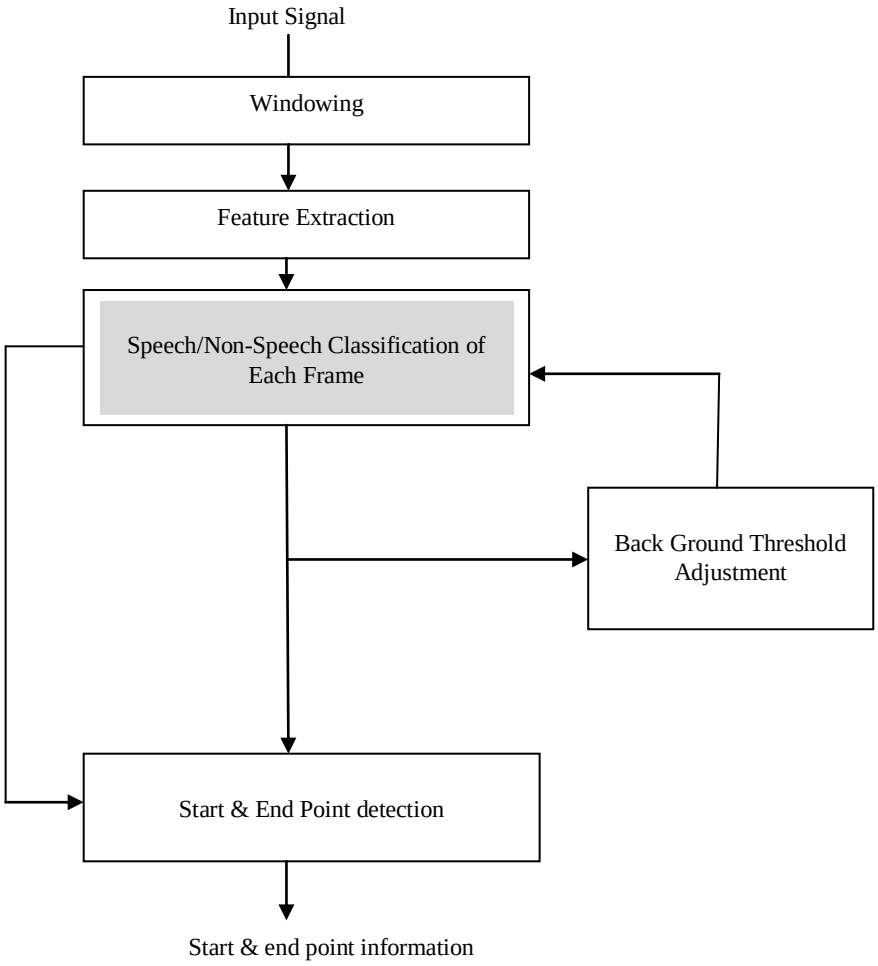
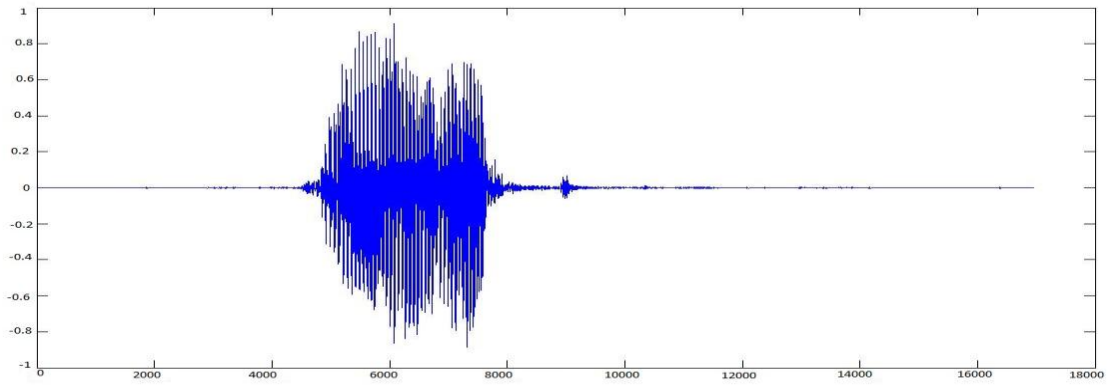


Figure 1 The Block diagram of general endpoint detection. [Ref. 16]



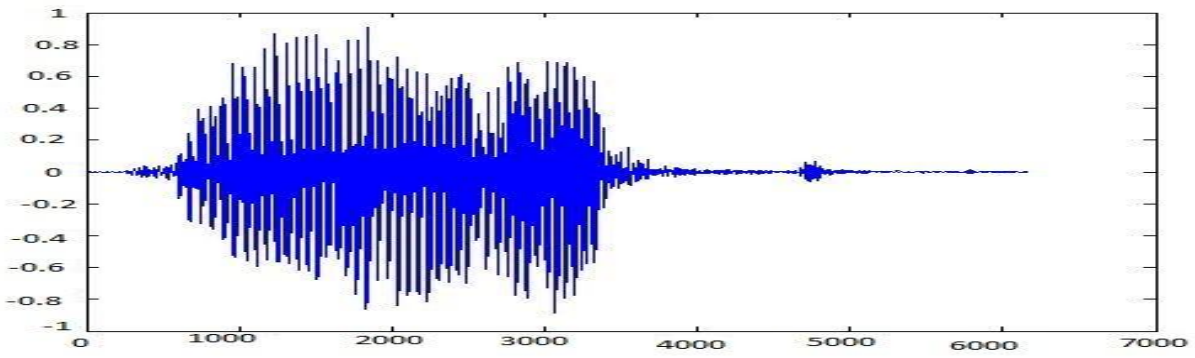


Figure-2 (a) Male speaker 1 in normal mode, waveform of isolated word "Bharat" in original form and after endpoint detection

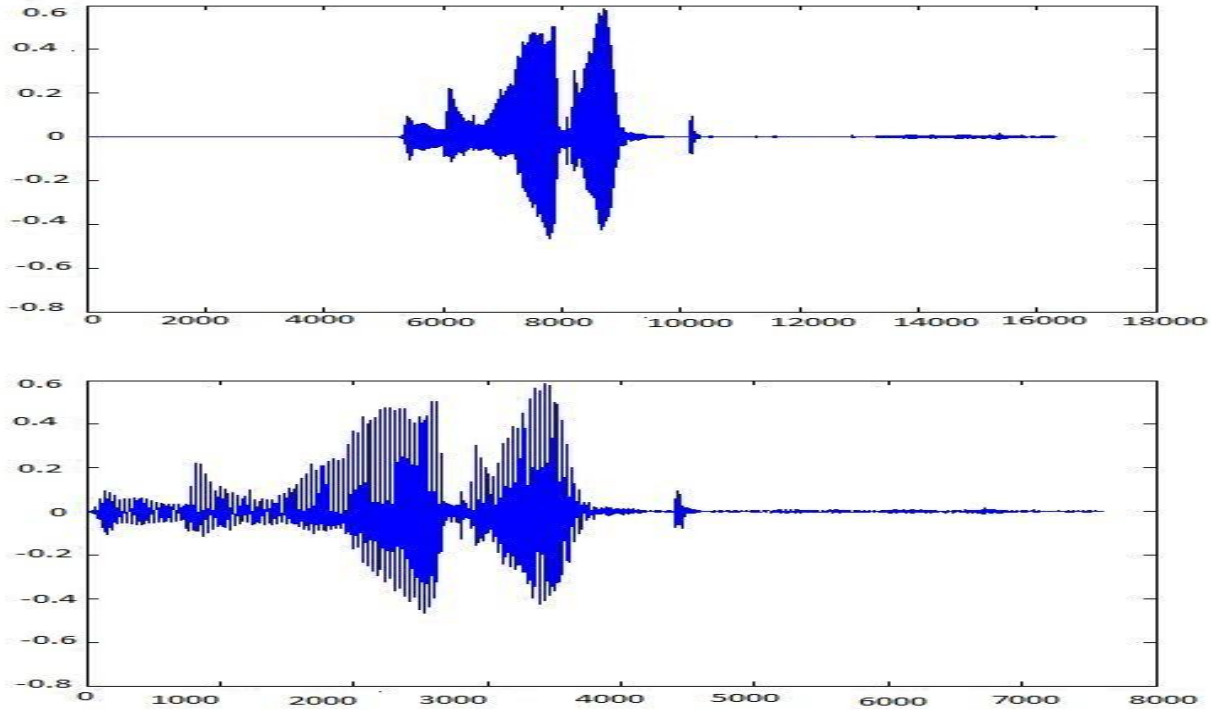
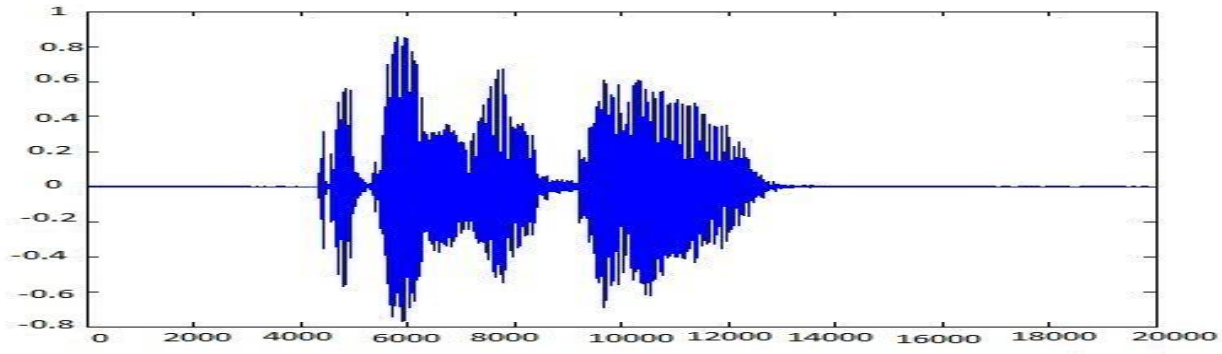


Figure-2 (b) Male speaker 1 in normal mode, waveform of isolated word "Bharat" in original form and after endpoint detection



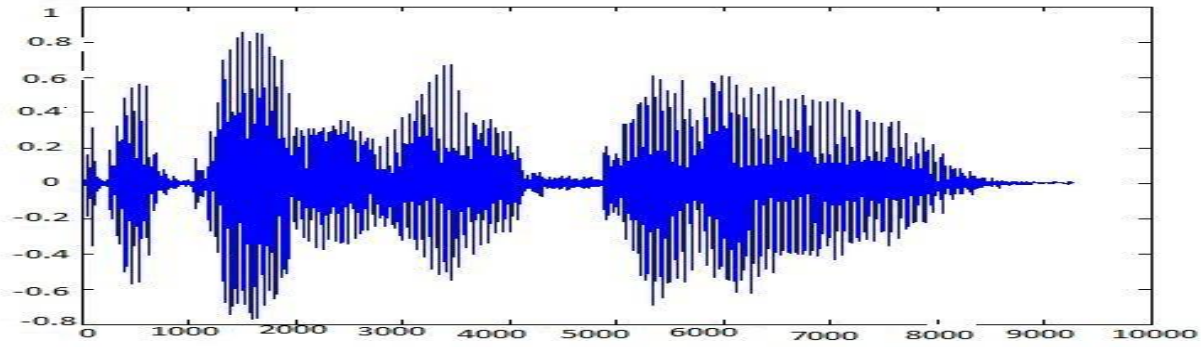


Figure-3 (a) Male speaker 1 in normal mode, waveform of SHHPW word "Kathin-Imtihaan" in original form and after endpoint detection

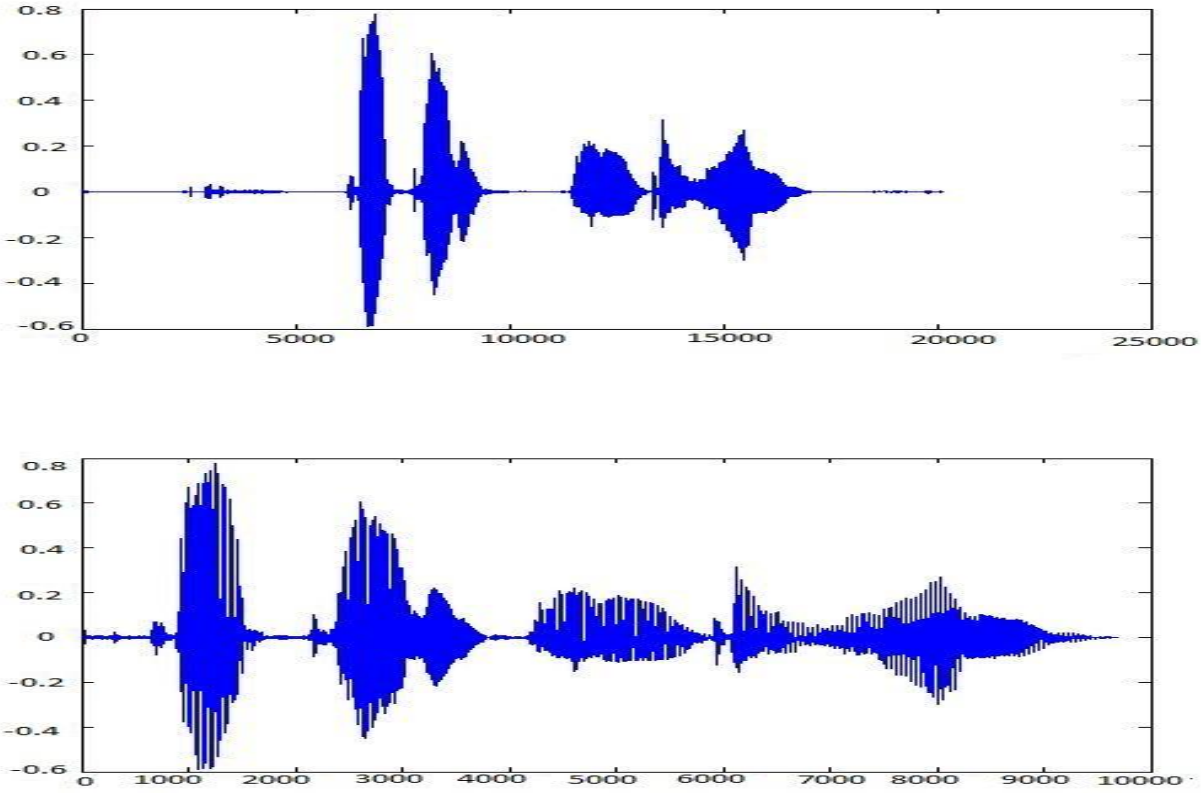
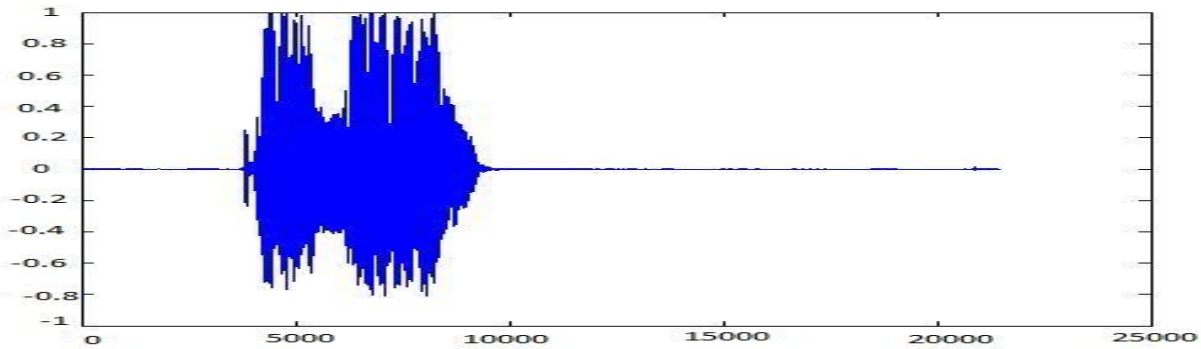


Figure-3 (b) Female speaker 1 in normal mode, waveform of SHHPW word "Kathin-Imtihaan" in original form and after endpoint detection



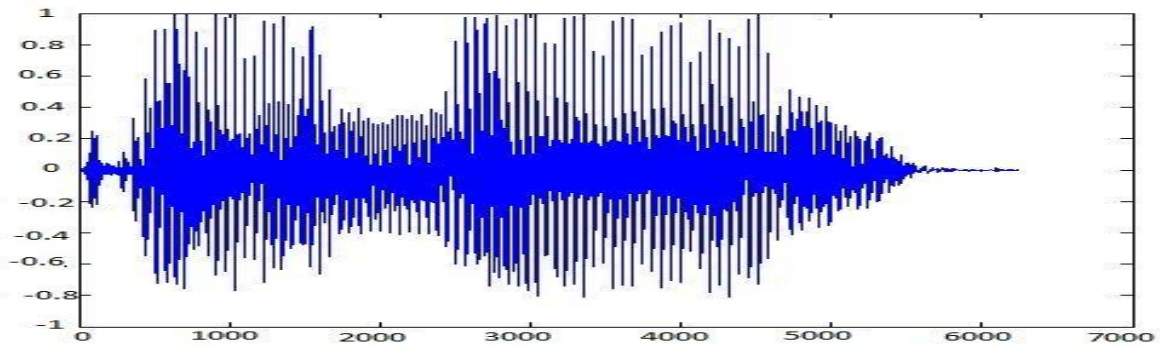


Figure-4 (a) Male speaker 1 in normal mode, waveform of SHPRW word "Kaam-Vaam" in original form and after endpoint detection

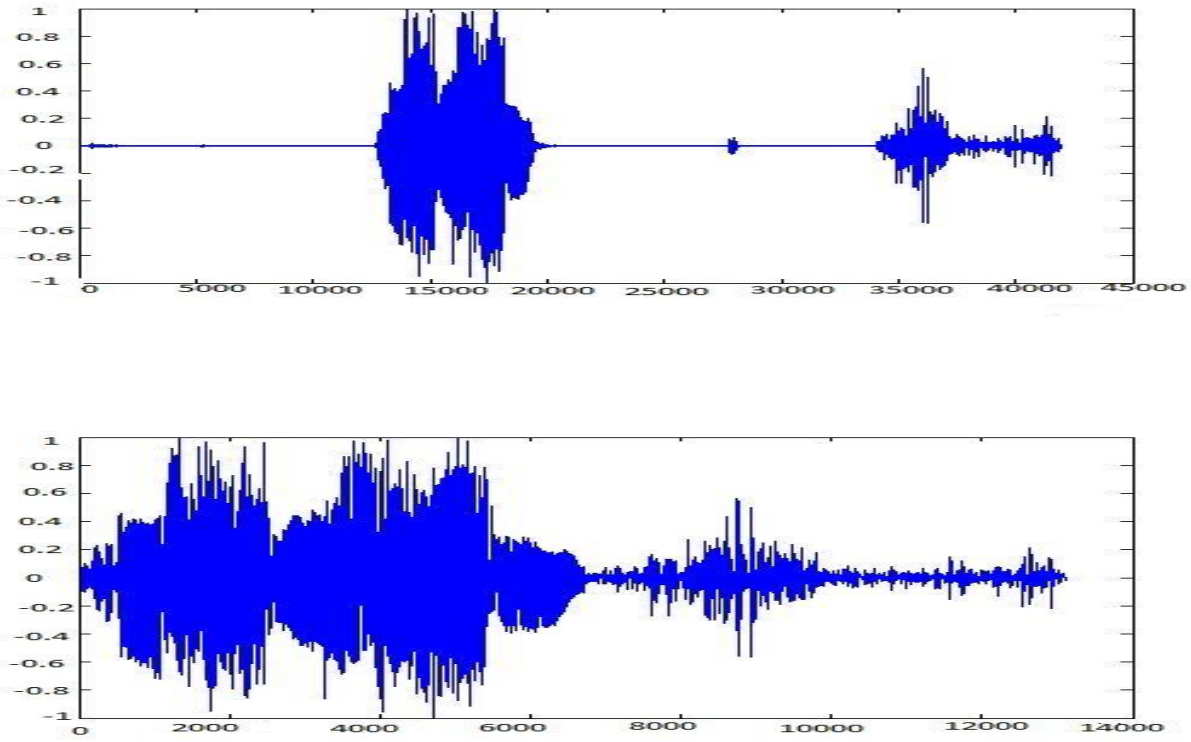


Figure-4 (b) Female speaker 1 in normal mode, waveform of SHPRW word "Kaam-Vaam" in original form and after endpoint detection

S.No	Parameters	WITHOUT UING END POINT DETECTION			USING END POINT DETECTION		
		WORDS			WORDS		
		ISOLATED	SHHPW	SHPRW	ISOLATED	SHHPW	SHPRW
a	Time Duration (in seconds)	2.12	2.48	2.68	0.77	1.16	0.78
2	Number of Samples	16000	19840	21440	6160	9280	6240
3	Root Mean Square(in Pascal)	0.1189	0.1102	0.124	0.1974	0.1612	0.231
4	Mean Power(intensity) in air(in db)	75.49	74.83	75.92	79.89	78.13	81.28

Table-1(a) Comparison of various parameters of male speakers for different words (average values)

S.No	Parameters	WITHOUT UING END POINT DETECTION			USING END POINT DETECTION		
		WORDS			WORDS		
		ISOLATED	SHHPW	SHPRW	ISOLATED	SHHPW	SHPRW
a	Time Duration (in seconds)	2.04	2.52	5.24	0.95	1.21	1.64
2	Number of Samples	16320	20160	41920	7600	13120	21440
3	Root Mean Square(in Pascal)	0.053	0.0653	0.1129	0.0784	0.0942	0.201
4	Mean Power(intensity) in air(in db)	68.56	70.28	75.04	71.87	73.47	80.08

Table-1(b) Comparison of various parameters of female speakers for different words (average values)

S.NO.	PARAMETERS	WORDS		
		ISOLATED	SHHPW	SHPRW
1	TD(in sec)	64.70(↓)	53.30(↓)	70.90(↓)
2	No. Of Samples	61.50(↓)	53.30(↓)	70.90(↓)
3	RMS (in pas.)	66.02(↑)	46.2(↑)	86.20(↑)
4	MPI(in db)	5.80(↑)	4.41(↑)	7.77(↑)

Table-2 (a) Increment (↑) and Decrement (↓) of all parameters after using end point detection in case of Male speakers (all values are in percentages)

S.NO.	PARAMETERS	WORDS		
		ISOLATED	SHHPW	SHPRW
1	TD(in sec)	53.50 (↓)	51.99 (↓)	68.71(↓)
2	No. Of Samples	53.50 (↓)	51.99 (↓)	68.71(↓)
3	RMS (in pas.)	47.92(↑)	44.27(↑)	78.03(↑)
4	MPI(in db)	4.82(↑)	4.53(↑)	6.71(↑)

Table-2 (b) Increment (↑) and Decrement (↓) of all parameters after using end point detection in case of Female speakers (all values are in percentages)

4. Conclusion

We describe the database development of three types of words and comparative analysis of these words using EPD algorithm. The result shows that average reduction is 70.90% and 70.90% in male while 68.71% and 68.71% in female in no. of samples and in time duration is the highest in case of SHPRW. The reduction in time duration and number of samples is greater than 50% in both male and female speakers so the processing time will be reduced and due to this response will be fast. Again by using result we can see that average enhancement of 86.20% and 7.77% in male while 78.03% and 6.71% in female in RMS and in MPI is the highest in SHPRW. So the Information content increases in these types of words and gap is also more unique comparable to Isolated and SHHPW.

5. References

1. D. K. Freeman, G. Cosier, C. B. Southcott, and I. Boyd, *"*The voice activity detector for the Pan-European digital cellular mobile telephone service*"*, in *Proc. Int. Conf. Acoust., Speech,*

Signal Processing, Glasgow, U.K., pp. 369–372, May 1989.
2. M.J. Hunt, M. Lennig and P. Mermelstein, "Experiments in syllable-based recognition of continuous speech", *Proc. IEEE Intl.Conf. Acoustics, Speech & Signal Processing*, Denver, pp.880-3.1983
3. L.R. Rabiner and B.H. Juang, B. Yegnanarayana, *Fundamentals of speech Recognition*, 1st edition, Pearson education in south Asia, 2009.
4. R. Cowie and R. R. Cornelius, "Describing the emotional states that are expressed in speech*", speech communication*, vol. 40, Apr. 2003, pp. 5-32.
5. L.R. Rabiner, R.W. Shafer, *Digital Processing of Speech Signals*, 3rd edition, Pearson education in south Asia, 2009.
6. V. Kumar, "A statistical approach towards the recognition of Hindi language words", *inria 00114544*, ver. 1, pp. 1–5, 2006.
7. H. Özer, "Signal detection and estimation in nonstationary background", *M.S. thesis*, Dept. Elect. Electron. Eng., Bas¸kent Univ., Ankara, Turkey, Aug1998.
8. L. Yang, "The expression and recognition of emotions through prosody", *in Proc. Int. Conf.Spoken Language Processing*, pp. 74–77, 2000.
9. Dinesh Kumar Rajoriya, R.S. Anand, R.P. Maheshwari, "Spoken Paired Word Pattern Classification Using Whole Word Template", *TECHNIA- Intl. J. of Computing Science and Communication Technologies*, vol.3, no.2, pp. 590-3, Jan. 2011.
10. R. Tucker, "Voice activity detection using a periodicity measure", *Proc.Inst. Elect. Eng.*, vol. 139, pp. 377–380, Aug. 1992.
11. K. Bullington and J. M. Fraser, "Engineering aspects of TASI", *Bell Syst.Tech. J.*, pp. 353–364, Mar. 1959.
12. G. Saha ,Sandipan Chakroborty,Suman Senapati, "A New Silence Removal and Endpoint Detection Algorithm for Speech and Speaker Recognition Applications".
13. Koji Kitayama, Masataka Goto, Katunobu Itou and Tetsunori Kobayashi, "Speech Starter: Noise-Robust Endpoint Detection by Using Filled Pauses", *Eurospeech, Geneva*, pp. 1237-1240, 2003.
14. S. E. Bou-Ghazale and K. Assaleh, "A robust endpoint detection of speech for noisy environments with application to automatic speech recognition", *in Proc. ICASSP2002*, vol. 4, 2002, pp. 3808–3811.
15. A. Martin, D. Charlet, and L. Mauuary, "Robust speech / non-speech detection using LDA applied to MFCC", *in Proc. ICASSP2001*, vol. 1, 2001, pp. 237–240.
16. Anand singh , Dr. Dinesh Kumar Rajoriya , Vikash Singh, "Database Development and Analysis of Spoken Hindi Hybrid Words Using Endpoint Detection" , *International Journal of Electronics and Computer Science Engineering*, vol1,no 3
17. Dinesh Kumar Rajoriya, R.S. Anand and R.P. Maheshwari, "Enhanced recognition rate of spoken Hindi paired word using probabilistic neural network approach" , *Int. J. Information and Communication Technology*, Vol. 3, No. 2, 2011
18. Q. Li, J. Zheng, A. Tsai and Q. Zhou, "Robust Endpoint Detection and Energy Normalization for Real-Time Speech and Speaker Recognition," *IEEE Transactions on Speech and Audio Processing*, vol.10, no.3, pp.146-157, March 2002.
19. Liang-Sheng Huang, Chung-Ho Yang, " A Novel Approach to Robust Speech Endpoint Detection in Car Environments", *In Proc. IEEE ICASSP-00*, pp.1751-pp.1754, 2000
20. Anand Singh, D. K. Rajoriya and Viaksh Singh, "Broad Acoustic Classification of Spoken Hindi Hybrid Paired Words using Artificial Neural Networks" , *International Journal of Computer Applications* (0975 – 8887) ,Volume 52– No.12, August 2012
21. Varsha Gupta, Mukul Pant, "An Approach To Describe Methods Of Front End Processing Of Speech Signal" , *International Journal of Scientific and Engineering Research*, Volume 4, Issue 3, Feb. 2013