International Journal of Computer Applications (0975 – 8887)
Volume 92 – No.15, April 2014

An Improved Endpoint Detection Algorithm using Bit Wise Approach for Isolated, Spoken Paired and Hindi Hybrid Paired Words

Taabish Gulzar
M.Tech. Student
Department of Electronics & Communication Engineering
D.I.T., Dehardun, India

Anand Singh
Assistant Professor
Department of Electronics & Communication Engineering
D.I.T., Dehradun, India

Sandip Vijay
Ph. D, Professor & HOD
Department of Electronics & Communication Engineering
D.I.T., Dehradun, India

ABSTRACT
Silence removal and endpoint detection using preset threshold values have already been used for locating the endpoints of an utterance. This paper presents a survey of bit by bit basis method for detecting the accurate endpoints of Isolated, Spoken Paired and Spoken Hindi Hybrid Paired words. Various parameters such as Number of samples, Time duration, Root Mean Square value and Mean Power (Intensity) in air are analyzed. The experiment results show that the proposed algorithm reduces the computational complexity.

General Terms
Pre-Processing, Recognition rate, Bit-Wise, Zero crossing rate, Speech Recognition.

Keywords
Bit wise analysis, endpoint detection, Paired words, Spoken Hindi Hybrid word.

1. INTRODUCTION
One of the most important ways of communication among humans is language and primary medium used for this communication is speech. Speech recognition via machines finds its applications in almost all sectors. Researchers are on their way for developing the systems that can provide better response to the users. The complicated assignment of speech processing has been divided in three relatively easier categories, (a) speech recognition: that allows the machines to understand the words, sentences, phrases spoken by humans, (b) Natural language processing: that allows the machines to understand the desire of humans, (c) speech synthesis: that allows the machines to respond back to users [1, 2]. Speech produced by same speaker different times, differ in the parameters containing within it such as energy, pitch, etc. For developing a robust speech system it is necessary to characterize the emotions present within a speech signal [3]. A unique category of words known as Spoken words are of different types such as short words, moderate words and long words [4].Spoken paired words have an edge over short and long words, as for as misrecognition rate, processing time, memory allocation is concerned [5]. Paired words overcome these problems as they contain moderate word length that requires less computation time and memory allocation. A gap of several milliseconds is always found in Paired words that play a vital role in recognition process, which depends on many factors, e.g. (length of vocal tract, air pressure through lungs, emotions of speaker, etc [6, 7]. In case of Hindi connected words only Hindi language is used for recognition purpose. However, to strengthen the speech code (gap between two words) two different languages are used. Hence a new category of words called Spoken Hindi Hybrid Paired words came in existence [8]. Hindi Hybrid words refer to the words in which one word is necessary from Hindi and other may be from any English, Urdu, Farsi, Sanskrit etc.

Pre-Processing of speech is essential in the applications where silence or background noise is entirely undesirable. The detection of the occurrence of speech segment embedded in different categories of non speech events and background noise is called endpoint detection, voice activity detection (VAD) or Silence removal. Silence removal techniques have been studied for several years. VAD finds its very first application in telephone transmission and switching system developed in Bell Labs, for the assignment of communication channel, where the channels idle time is detected by the presence of users speech and consequently an unused channel is assigned when speech is detected, providing more customer services [9]. The falsely detected endpoints of an utterances results in at least two negative effects [10]:

1. Recognition errors are introduced due to the inaccurate boundaries of the speech.
2. Increases the computation if inaccurate boundaries are detected.

Two broadly accepted methods namely Short Time Energy and Zero Crossing Rate (ZCR) have been used since a long time for removing the background noise from speech [11, 12]. Unfortunately these algorithms have some problems regarding setting the thresholds as ad hoc basis. This value of threshold is often assumed to be fixed or computed in voice-inactive (silence) intervals of a speech signal [13].

A newer promising approach involves the use of Bit-Wise analysis to locate the accurate endpoints where the concept of setting the threshold as ad hoc basis is avoided. The proposed method reduces the computational complexity and at the same time it does not make use of ZCR, which sometimes is not feasible. Another advantage of using this method is that there is no need for the calculation of background noise. Hence it is efficient for the environment, where there is strong noise at the beginning or ending of the speech or when the speech artifacts such as breath, mouth and lip clicks are likely to happen which in turn locate the false boundaries.

The rest of the paper is organized as follows: In Section II, we will introduce the theoretical background. Section III briefly describes the Bit-wise algorithm for evaluating the endpoints of the utterances. Section IV presents the database preparation and experimental results. Finally, we will summary our findings in section IV.

2. THEORETICAL BACKGROUND

Speech signal is a slowly time varying signal, when seen over a relatively short interval of time, and its characteristics are quite stationary [14]. In order to meet the requirements that include computational accuracy, complexity, response time etc different applications make use of different algorithms. These applications include those which are based on energy threshold, pitch detection, ZCR [15, 16, 17, 18].The end points of speech are usually obscured by speaker generated artifacts such as clicks or by dial tone. Long-distance telephone transmission channels may also introduce these types of artifacts and some background noise [19]. Conventional short-time or spectral energy or ZCR based endpoint detection algorithms are usually susceptible to speech artifacts such as breath, mouth and lip clicks and break down easily in the presence of noise. These classical energy threshold methods i.e. energy and ZCR methods, the threshold value need to be recalculated at each and every silence (voice-inactive) segment [20]. And in case, the noise is non-stationary, these methods fail to track the exact value of thresholds, resulting in falsely detected endpoints. In some of the applications which may include speaker verification, name dialing, speech control etc, where the speech (voice-active) part of the signal is sometimes less (e.g. less than 2 s) and the recognition process can be done within 1 s or even less, that are usually provided by embedded systems, such as wireless phones or portable devices; or in multi-user systems, such as speaker verification system for several users, usually a low computational complexity for low cost or for faster response of the system is required [21]. One solution for the abovementioned cases is to make use of an accurate endpoint detection algorithm to remove all silence (voice-inactive) part. Intrusion of the proposed Bit-wise method uniquely defines the threshold first instance.

3. PROPOSED ALGORITHM

In this section, an algorithm for endpoint detection that makes use of bit by bit comparison of the signal is presented and meets the following requirements: accurate detection of endpoints, low computational complexity, fast response time and simple steps of implementations.

Step 1: Removing zeroes from the signal:

The adjacent bits of the speech signal are checked and zeroes i.e. the silence (voice-inactive) intervals are removed.

Step 2: Calculation of threshold:

The signal is then transposed. If μ and σ are the values of mean and standard deviation respectively then mathematically,

$$\sigma = \sqrt{1/N\sum_{i=1}^{N}(\chi_i - \mu)^2} \qquad (1)$$

$$\mu = \sum_{i=1}^{N} \chi_i \qquad (2)$$

Where, x_i = observed values of the sample item,

N= size of the sample.

Step 3: Finding the start point of signal:

Once a threshold has been estimated, anything above this threshold is considered to be speech and anything below this optimum point is either noise or silence.

$$\text{Signal} = \begin{cases} \text{speech}, & \text{if signal (i)} \geq \text{threshold}, \ i=1,2,\ldots n \\ \text{silence ,otherwise} & \end{cases} \qquad (3)$$

If the difference between the successive bits of the signal is greater than the threshold value calculated, it is the start point of the signal and the signal is stored in that case.

Step 4: Finding the endpoint of the signal:

The threshold calculated in Step 2 is taken for estimating the endpoint of the signal. Now, if the difference between any of the bit and its preceding one is greater than the threshold calculated, this point is treated as the endpoint of the signal, and is stored.

Step 5: The resulting signal is the accurately endpoint detected signal.

3.1 Set Up

Step (1):- Initialize required variables

Step (2):- Read file, find length of the signal, Remove one channel.

Algorithm for endpoint detection:-

Step (3):- Remove Zeroes from the signal.

for i ← 1 to r

if ← y(i) not 0

yy(k) ← y(i)

k ← k+1

end if

end for

Step (4):- Transpose signal to find end point.

Step (5):- Calculate threshold (Thold) .

for i ← 1 to r

m = m + (yy(i) – mean(yy))^2 ;

end for

*Thold = sqrt(1/r * m);*

Step (6):- Find the start point.

for i ← 1 to r

if yy(i) –yy(i+1) greater than Thold

break;

end if

end for

Step (7):- Store remaining signal.

yyy ← yy(1 to r, 1)

Step (8):- Calculate threshold as calculated in Step (5).

Step (9):-Find the endpoint.

for i ← r to 2 step -1

if yyy(i-1) – yyy(i) greater than Thold

break;

end if

end for

Step (10):- Store the remaining signal as EPDsignal.

EPDsignal ← yyy(1 to r)

Step (11):- The resultant signal is the endpoint detected signal

4. DATABASE DEVELOPMENT AND EXPERIMENTAL RESULTS.

Thirty different words are used for database, out of which ten words are Isolated, ten Spoken paired, and the remaining ten spoken Hindi hybrid words. Five female and five male speakers are taken for recording session, thus making a total of three hundred utterances. Stereo headset H 250 with frequency response of 20 Hz-20 KHz was used. A sampling rate of 16000 Hz was used for all data. For some of the recordings there is some background or some lip and mouth clicks or breath noise.

The Pre-processing, endpoint detection technique was implemented and verified in MATLAB and PRAAT software's respectively. To evaluate performance, we visually compare the locations of the beginning and ending points using Bit wise and preset threshold approach. The detected endpoint signal is verified in PRAAT software for analyzing various parameters i.e. Number of samples (NOS), Time duration (TD) in seconds, Root mean square (RMS) in Pascal and Mean power intensity in air (MI) in dB. Figure 1 to figure 18 compares the waveforms of Isolated word (TEEN), Spoken Paired word (HUM-TUM) and Hybrid word (KHAS-AAMI) in their original form and endpoint detected form using Bit wise and preset threshold approaches for both female and male speakers respectively. Table 1 to Table 6 shows the average value of various parameters i.e. NOS, TD, RMS, MI for Isolated, Paired and Hybrid words for both female and male speakers respectively. Table 7 to Table 12 gives the increment and decrement in the abovementioned parameters for both female and male speakers respectively using Bit wise approach. Proposed algorithm performs well than conventional threshold method.

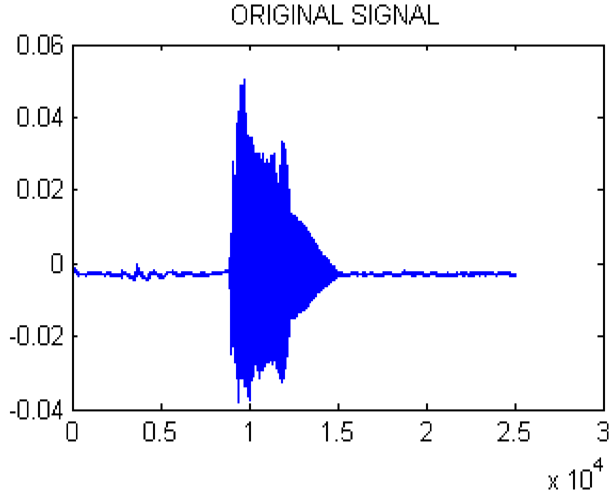

Figure 1: Waveform of isolated word (TEEN) by female speaker in the original form.

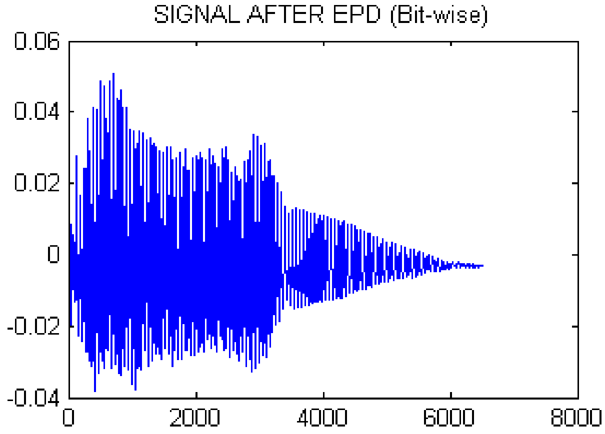

Figure 2: Waveform of isolated word (TEEN) by female speaker after endpoint detection (Bit wise).

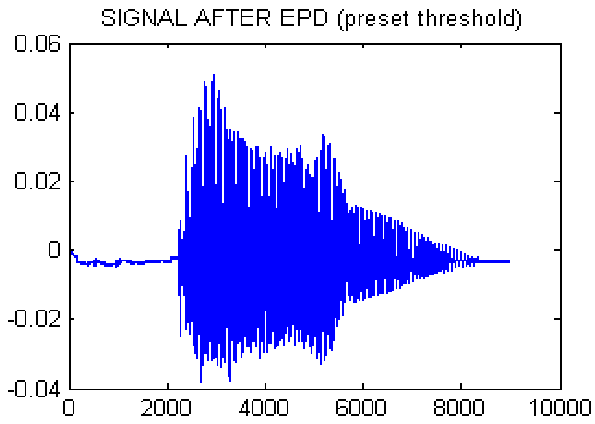

Figure 3: Waveform of isolated word (TEEN) by female speaker after endpoint detection (preset Threshold).

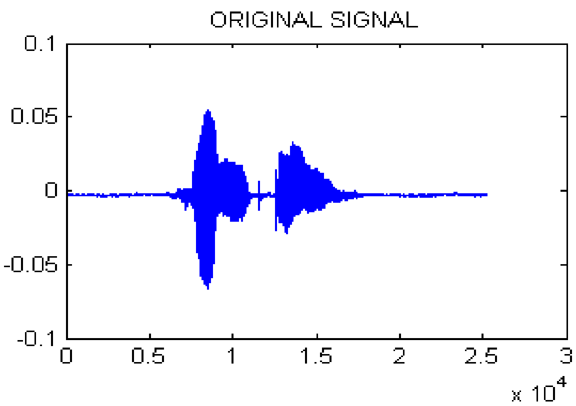

Figure 4: Waveform of paired word (HUM-TUM) by female speaker in original form.

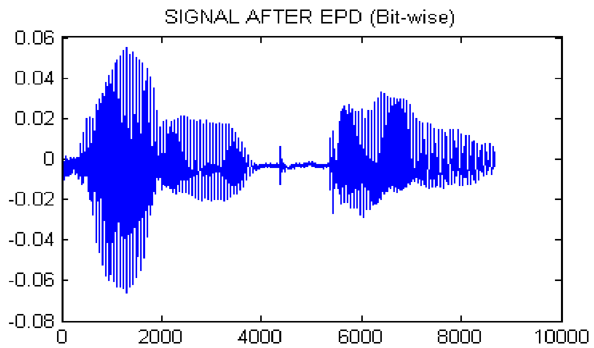

Figure 5: Waveform of paired word (HUM-TUM) by female speaker after endpoint detection (Bitwise).

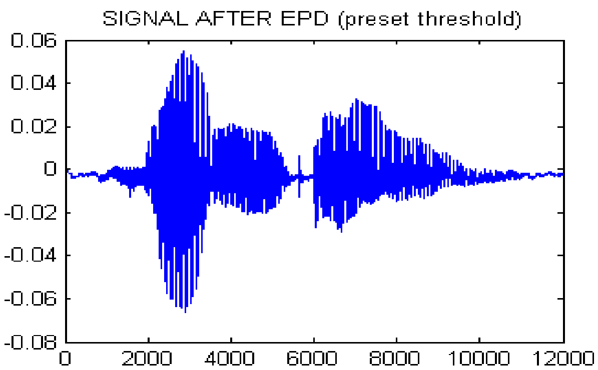

Figure 6: Waveform of paired word (HUM-TUM) by female speaker after endpoint detection (preset Threshold).

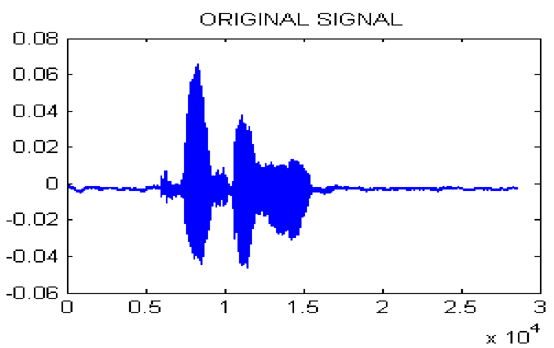

Figure 7: Waveform of Hybrid word (KHAS-AADMI) by female speaker in original form.

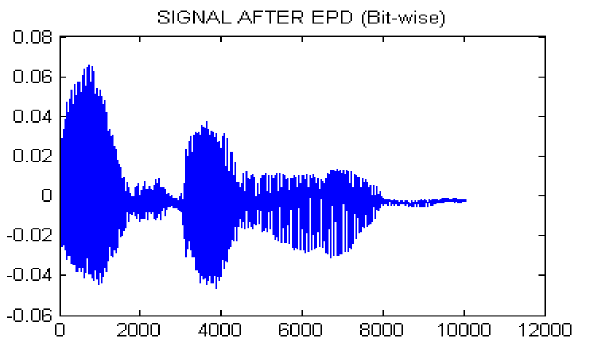

Figure 8: Waveform of Hybrid word (KHAS-AADMI) by female speaker after endpoint detection (Bitwise).

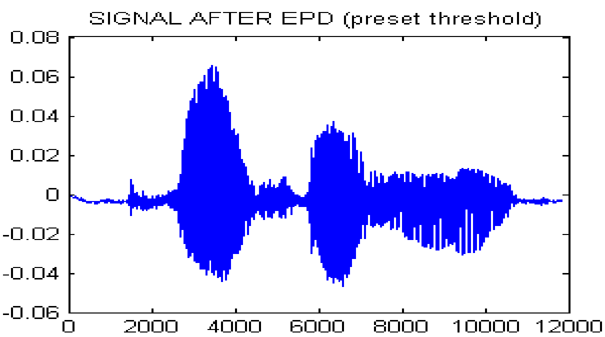

Figure9: Waveform of Hybrid word (KHASS-AADMI) by female Speaker after endpoint detection (preset Threshold).

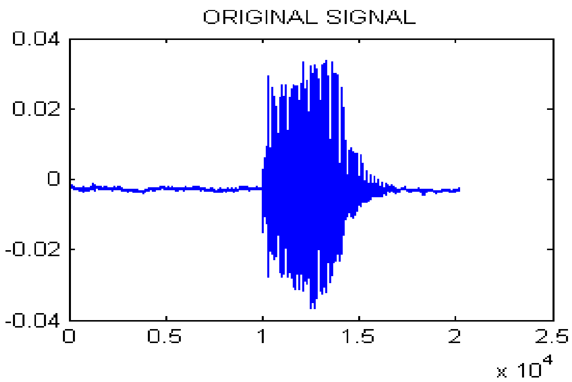

Figure 10: Waveform of isolated word (TEEN) by male speaker in original form.

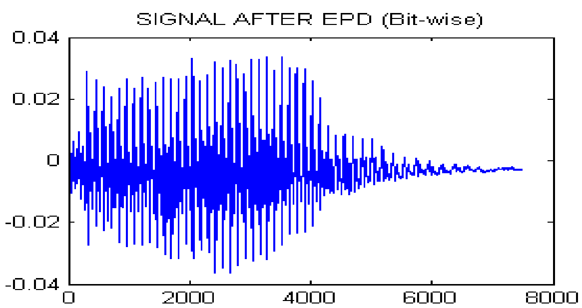

Figure- 11 Waveform of isolated word (TEEN) by male speaker after endpoint detection (Bit wise).

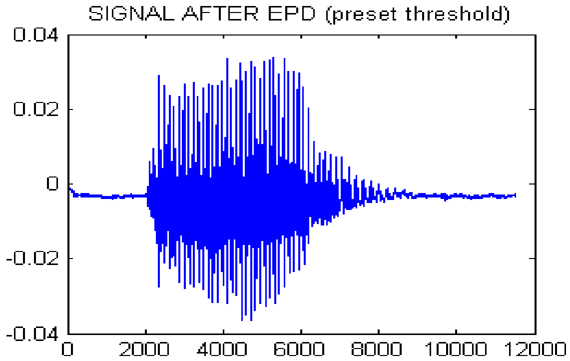

Figure 12: Waveform of isolated word (TEEN) by male speaker after endpoint detection (preset Threshold).

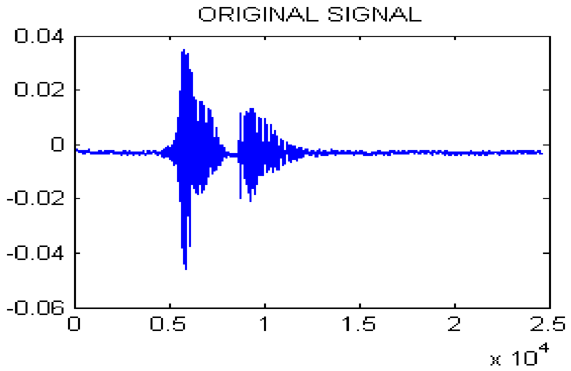

Figure 13: Waveform of paired word (HUM-TUM) by male Speaker in original form.

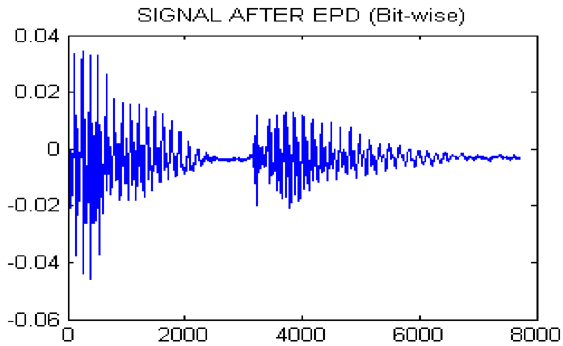

Figure 14: Waveform of paired word (HUM-TUM) by male speaker after endpoint detection (Bit wise).

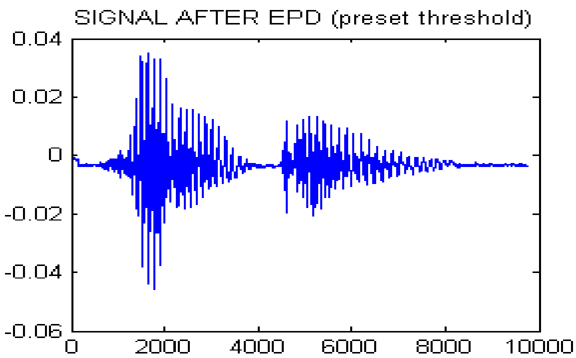

Figure 15: Waveform of paired word (HUM-TUM) by male speaker after endpoint detection (preset Threshold).

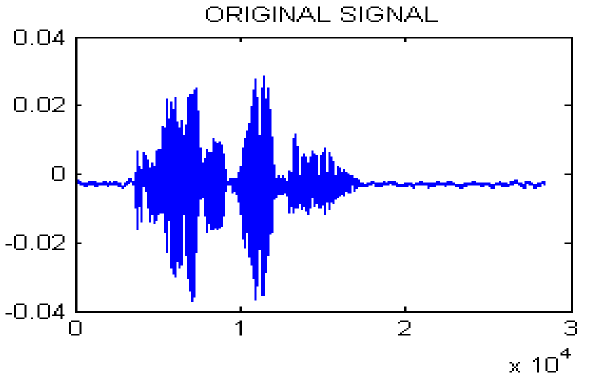

Figure 16: Waveform of Hybrid word (KHAS-AADMI) by male Speaker in original form.

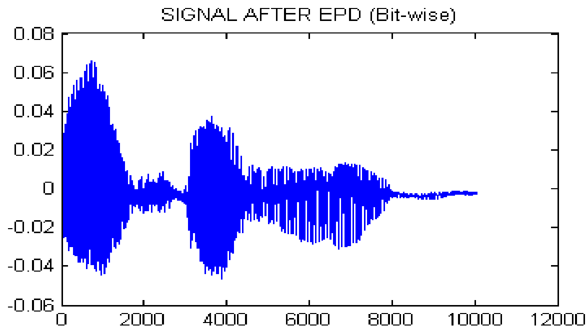

Figure 17: Waveform of Hybrid word (KHAS-AADMI) by male Speaker after endpoint detection (Bit wise).

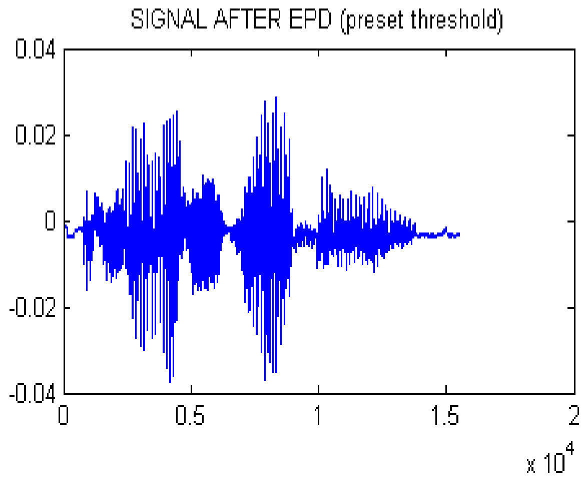

Figure 18: Waveform of Hybrid word (KHAS-AADMI) by male speaker after endpoint detection (preset Threshold).

Table 1. Comparison of various parameters of female speakers for isolated words (average values).

WORDS	EPD USING PRESET THRESHOLD METHOD				EPD USING BIT BY BIT ANALYSIS			
	NOS	TD	RMS	MI	NOS	TD	RMS	MI
SHOONYA	12416	0.776	0.016394	57.018	9054.8	0.565928	0.018936	58.25
EK	11296	0.706	0.03593	58.742	7438	0.464878	0.051728	60.192
DO	7968	0.498	0.01031	54.018	5455.8	0.34099	0.01207	55.368
TEEN	9184	0.574	0.00869	52.532	6529	0.408064	0.009916	53.718
CHAR	9824	0.614	0.00893	54.526	7931.2	0.495702	0.011942	55.346
PAANCH	10272	0.642	0.015944	57.626	8004.6	0.520288	0.017846	58.58
CHEH	9920	0.62	0.012438	55.56	8341.6	0.521386	0.013536	56.274
SAATH	9408	0.588	0.01603	57.34	6264.8	0.391552	0.01949	58.954
AATH	9120	0.57	0.014774	57.118	6329	0.395564	0.017426	58.568
NAU	8320	0.52	0.011062	54.49	5767	0.360434	0.012452	55.524

Table 2. Comparison of various parameters of female speakers for paired words (average values).

WORDS	EPD USING PRESET THRESHOLD METHOD				EPD USING BIT BY BIT ANALYSIS			
	NOS	TD	RMS	MI	NOS	TD	RMS	MI
ANGE- PINCHE	14624	0.914	0.011136	54.32	13420.4	0.838768	0.011574	54.65
UPPER-NICHE	14272	0.88	0.012386	55.05	12081.6	0.755102	0.0131378	55.59
DIN-RAAT	13280	0.83	0.024152	58.57	10093.2	0.630824	0.028846	59.63
HUM-TUM	11712	0.732	0.013649	56.24	9087	0.567936	0.0156	57.292
DHAN-DAULAT	17728	1.108	0.014132	56.7	12559.4	0.784962	0.01669	57.986
JINA-MARNA	15648	0.978	0.012938	55.564	12958	0.809876	0.013954	56.328
UTHANA-BAITHANA	17216	1.076	0.0234118	59.482	12519	0.78244	0.027726	60.684
RANG-BERANG	17216	1.076	0.014266	56.626	14617.7	0.913596	0.015584	57.286
KHANA-PINA	16608	1.038	0.011266	54.48	15104.4	0.944024	0.011922	54.814
YANHA-WANHA	15456	0.966	0.01326	55.432	11440.6	0.71504	0.014922	56.528

Table 3. Comparison of various parameters of female speakers for Hybrid words (average values).

WORDS	EPD USING PRESET THRESHOLD METHOD				EPD USING BIT BY BIT ANALYSIS			
	NOS	TD	RMS	MI	NOS	TD	RMS	MI
KAALA-SAYA	16896	1.056	0.033308	58.164	13526.4	0.8454	0.039328	59.016
KHAS-AADMI	16768	1.048	0.021374	57.36	10828.6	0.676788	0.025694	58.996
BADIYA-ITEM	16160	1.01	0.024438	58.412	12624.6	0.789039	0.027958	59.226
DOUBLE-ROTI	13728	0.858	0.012306	55.02	12428.8	0.776682	0.012826	55.364
PURANA-ZAMANA	16768	1.048	0.016336	57.114	13209.4	0.825592	0.017902	57.924
BADA-EHSAAN	16325	1.022	0.021038	58.208	11892.4	0.743276	0.024334	59.328
BADIYA-JOKE	14816	0.926	0.029444	58.176	11407	0.712938	0.032022	59.072
PURANI-JEANS	18112	1.132	0.022064	57.468	14493.6	0.905364	0.024588	58.272
BADIYA-SAWAAL	16704	1.044	0.016128	55.81	13876.4	0.867274	0.017196	56.324
TERI –IBADAT	17056	1.066	0.017752	56.474	13583.2	0.848952	0.019832	57.348

Table 4. Comparison of various parameters of male speakers for isolated words (average values).

WORDS	EPD USING PRESET THRESHOLD METHOD				EPD USING BIT BY BIT ANALYSIS			
	NOS	TD	RMS	MI	NOS	TD	RMS	MI
SHOONYA	10484	0.664	0.012568	54.172	8717.4	0.544728	0.014323	54.892
EK	8288	0.518	0.026034	59.48	6540.8	0.408802	0.026472	60.174
DO	9280	0.58	0.065734	60.164	5892	0.368254	0.089972	61.784
TEEN	9152	0.572	0.009406	53.136	6648.2	0.415512	0.010664	54.252
CHAR	9120	0.57	0.012812	54.848	7666.6	0.479164	0.013988	55.426
PAANCH	11936	0.746	0.045964	62.266	7995.8	0.499738	0.0592178	63.866
CHEH	10720	0.67	0.045352	63.712	7268.4	0.454725	0.058856	65.154
SAATH	10176	0.636	0.054674	63.762	7059.6	0.441224	0.064128	65.21
AATH	10272	0.642	0.070688	61.198	7495	0.468438	0.082612	62.466
NAU	9952	0.622	0.064746	60.828	5598.8	0.348828	0.10009	62.788

Table 5. Comparison of various parameters of male speakers for paired words (average values).

WORDS	EPD USING PRESET THRESHOLD METHOD				EPD USING BIT BY BIT ANALYSIS			
	NOS	TD	RMS	MI	NOS	TD	RMS	MI
ANGE- PINCHE	15008	0.938	0.010682	53.692	13066.6	0.816666	0.011284	54.136
UPPER-NICHE	13824	0.864	0.013926	55.442	10278.8	0.642454	0.014712	55.988
DIN-RAAT	13024	0.814	0.013106	54.938	9949.8	0.621864	0.01481	55.896
HUM-TUM	9792	0.612	0.006716	50.256	8374.6	0.523414	0.007112	50.708
DHAN-DAULAT	13728	0.858	0.010736	53.908	11509.2	0.719326	0.01145	54.572
JINA-MARNA	15968	0.998	0.008186	51.625	11851	0.740688	0.009456	52.74
UTHANA-BAITHANA	14016	0.874	0.010338	52.672	12563	0.78519	0.010608	52.904
RANG-BERANG	13600	0.85	0.009114	52.076	10962.5	0.685328	0.009974	52.806
KHANA-PINA	15520	0.97	0.008002	50.6	12713	0.794562	0.008592	51.166
YANHA-WANHA	13760	0.86	0.00944	53.04	11078.6	0.692414	0.010394	53.808

Table 6. Comparison of various parameters of male speakers for Hybrid words (average values).

WORDS	EPD USING PRESET THRESHOLD METHOD				EPD USING BIT BY BIT ANALYSIS			
	NOS	TD	RMS	MI	NOS	TD	RMS	MI
KAALA-SAYA	15776	0.986	0.010098	53.098	12069.6	0.754352	0.011712	54.11
KHAS-AADMI	15840	0.99	0.00885	51.576	12889.6	0.80574	0.009778	52.24
BADIYA-ITEM	14048	0.878	0.01022	53.482	11788	0.7367516	0.0112762	54.16
DOUBLE-ROTI	15040	0.94	0.00691	50.422	11770.4	0.735652	0.00766	51.226
PURANA-ZAMANA	16928	1.058	0.007182	50.974	14181.2	0.886328	0.007662	51.54
BADA-EHSAAN	15974	1.004	0.008236	51.798	12619.8	0.788736	0.009038	52.576
BADIYA-JOKE	13504	0.844	0.0104	53.406	10469.8	0.654364	0.011834	54.318
PURANI-JEANS	16704	1.044	0.00648	49.804	14095.8	0.880988	0.006864	50.242
BADIYA-SAWAAL	14016	0.876	0.006332	49.804	12139.6	0.758926	0.006598	50.162
TERI -IBADAT	15136	0.946	0.008112	51.798	13577.2	0.848576	0.008514	52.176

Table 7. Increment (↑) and Decrement (↓) of all parameters in case of female speakers for Isolated words (in percentage).

WORDS	NOS	TD	RMS	MI
SHOONYA	27.0712(↓)	27.07113(↓)	15.50567(↑)	2.16072(↑)
EK	34.15368(↓)	34.15326(↓)	43.96883(↑)	2.46842(↑)
DO	31.52861(↓)	31.52811(↓)	17.07081(↑)	2.49917(↑)
TEEN	28.90987(↓)	28.90871(↓)	14.10817(↑)	2.25767(↑)
CHAR	19.26710(↓)	19.26645(↓)	33.72900(↑)	1.50387(↑)
PAANCH	22.07359(↓)	22.07352(↓)	11.92925(↑)	1.65550(↑)
CHEH	15.91129(↓)	15.90548(↓)	8.82779(↑)	1.28509(↑)
SAATH	33.40986(↓)	33.40982(↓)	21.58453(↑)	2.81479(↑)
AATH	30.60307(↓)	30.60281(↓)	17.95045(↑)	2.52860(↑)
NAU	30.68509(↓)	30.68577(↓)	12.56554(↑)	1.89759(↑)

Table 8. Increment (↑) and Decrement (↓) of all parameters in case of female speakers for Paired words (in percentage).

WORDS	NOS	TD	RMS	MI
ANGE- PINCHE	8.23031(↓)	8.23107(↓)	3.93319(↑)	0.60751(↑)
UPPER-NICHE	15.34753(↓)	14.19295(↓)	6.06976(↑)	0.98093(↑)
DIN-RAAT	23.99699(↓)	23.99711(↓)	19.43524(↑)	1.80980(↑)
HUM-TUM	22.41291(↓)	22.41311(↓)	13.91850(↑)	1.87055(↑)
DHAN-DAULAT	29.15501(↓)	29.15505(↓)	18.10076(↑)	2.26807(↑)
JINA-MARNA	17.19069(↓)	17.01959(↓)	7.85284(↑)	1.37499(↑)
UTHANA-BAITHANA	27.28276(↓)	27.28253(↓)	18.42746(↑)	2.02078(↑)
RANG-BERANG	15.09294(↓)	15.09330(↓)	9.23875(↑)	1.16554(↑)
KHANA-PINA	9.05347(↓)	9.05356(↓)	5.82283(↑)	0.61307(↑)
YANHA-WANHA	25.97955(↓)	25.97929(↓)	12.53394(↑)	1.97719(↑)

Table 9. Increment(↑) and Decrement (↓) of all parameters in case of female speakers for Hybrid words (in percentage).

WORDS	NOS	TD	RMS	MI
KAALA-SAYA	19.94318(↓)	19.94318(↓)	18.07374(↑)	1.46482(↑)
KHAS-AADMI	35.42104(↓)	35.42099(↓)	20.21147(↑)	2.85216(↑)
BADIYA-ITEM	21.87747(↓)	21.87733(↓)	14.40379(↑)	1.39355(↑)
DOUBLE-ROTI	9.46387(↓)	9.47762(↓)	4.22588(↑)	0.62523(↑)
PURANA-ZAMANA	21.22257(↓)	21.22214(↓)	9.58619(↑)	1.41822(↑)
BADA-EHSAAN	27.27250(↓)	27.27241(↓)	15.66689(↑)	1.92413(↑)
BADIYA-JOKE	23.00891(↓)	23.00886(↓)	8.75560(↑)	1.54015(↑)
PURANI-JEANS	19.97792(↓)	20.02085(↓)	11.43945(↑)	1.39904(↑)
BADIYA-SAWAAL	16.92768(↓)	16.92778(↓)	6.62202(↑)	0.95323(↑)
TERI –IBADAT	20.36116(↓)	20.36098(↓)	11.71699(↑)	1.54761(↑)

Table 10. Increment(↑) and Decrement (↓) of all parameters in case of male speakers for Isolated words (in percentage).

WORDS	NOS	TD	RMS	MI
SHOONYA	16.87905(↓)	17.96265(↓)	13.96404(↑)	1.32909(↑)
EK	21.08108(↓)	21.08069(↓)	1.69009(↑)	1.27234(↑)
DO	36.50862(↓)	36.50793(↓)	36.87285(↑)	2.69264(↑)
TEEN	27.35794(↓)	27.35804(↓)	13.37444(↑)	2.10027(↑)
CHAR	15.93640(↓)	15.93641(↓)	9.17889(↑)	1.05382(↑)
PAANCH	33.01106(↓)	33.01099(↓)	28.83561(↑)	2.56962(↑)
CHEH	32.19776(↓)	32.13059(↓)	29.77578(↑)	2.26331(↑)
SAATH	30.625(↓)	30.62516(↓)	17.29158(↑)	2.27095(↑)
AATH	27.03466(↓)	27.03458(↓)	16.86849(↑)	2.07196(↑)
NAU	43.74196(↓)	43.74148(↓)	54.5887(↑)	3.22220(↑)

Table 11. Increment (↑) and Decrement (↓) of all parameters in case of male speakers for Paired words (in percentage).

WORDS	NOS	TD	RMS	MI
ANGE- PINCHE	12.93574(↓)	12.93539(↓)	5.635652(↑)	0.82694(↑)
UPPER-NICHE	25.64525(↓)	25.64189(↓)	5.64412(↑)	0.98481(↑)
DIN-RAAT	23.60412(↓)	23.60393(↓)	13.00168(↑)	1.74378(↑)
HUM-TUM	14.47508(↓)	14.47484(↓)	5.89637(↑)	0.89939(↑)
DHAN-DAULAT	16.16259(↓)	16.16247(↓)	6.65052(↑)	1.23173(↑)
JINA-MARNA	25.78282(↓)	25.78277(↓)	15.51429(↑)	2.10640(↑)
UTHANA-BAITHANA	11.08019(↓)	10.16133(↓)	2.61172(↑)	0.44046(↑)
RANG-BERANG	19.39338(↓)	19.37318(↓)	9.43603(↑)	1.40179(↑)
KHANA-PINA	18.08634(↓)	18.08639(↓)	7.37316(↑)	2.09486(↑)
YANHA-WANHA	19.48692(↓)	19.48674(↓)	10.10593(↑)	1.44796(↑)

Table 12. Increment (↑) and Decrement (↓) of all parameters in case of male speakers for Hybrid words (in percentage).

WORDS	NOS	TD	RMS	MI
KAALA-SAYA	23.49391(↓)	23.49371(↓)	15.98336(↑)	1.90591(↑)
KHAS-AADMI	18.62626(↓)	18.61212(↓)	10.48588(↑)	1.28742(↑)
BADIYA-ITEM	16.08769(↓)	16.08762(↓)	10.33464(↑)	1.26772(↑)
DOUBLE-ROTI	21.73936(↓)	21.73915(↓)	10.85384(↑)	1.59454(↑)
PURANA-ZAMANA	16.22637(↓)	16.22609(↓)	6.68338(↑)	1.11037(↑)
BADA-EHSAAN	20.99787(↓)	21.44064(↓)	9.73774(↑)	1.50199(↑)
BADIYA-JOKE	22.46889(↓)	22.46872(↓)	13.78846(↑)	1.70767(↑)
PURANI-JEANS	15.61422(↓)	15.61418(↓)	5.92593(↑)	0.87945(↑)
BADIYA-SAWAAL	13.38356(↓)	13.36461(↓)	4.20088(↑)	0.71882(↑)
TERI –IBADAT	10.29863(↓)	10.29852(↓)	4.95562(↑)	0.71238(↑)

5. CONCLUSION

A new Bit wise approach for detecting the accurate endpoints of a speech signal is presented. Complete details regarding the implementation of the algorithm have been provided. The experimental results demonstrate that the proposed algorithm can be used to enhance the performance of endpoint detection algorithms. The result shows average reduction of 27.36% and 27.36% for Isolated words, 19.37% and 19.26% for Paired words and 21.55% and 21.55% for Hybrid words in NOS and TD, while average enhancement of 19.72% and 2.11% for Isolated words, 11.53% and 1.47% for Paired words and 12.07% and 1.51% for Hybrid words in RMS and MI values for female speakers respectively. For male speakers an average reduction of 28.44% and 28 .54% for Isolated words, 18.67% and 18.57% for Paired words and 17.89% and 17.93% for Hybrid words in NOS and TD, while as an average enhancement of 22.24% and 2.08% for Isolated words, 8.19% and 1.32% for Paired words and 9.29% and 1.27% for Hybrid words in RMS and MI is seen respectively. Due to reduction in NOS and TD, processing time will be reduced and enhancement in RMS and MI values increases the information content. The proposed Bit-wise endpoint detection method may lead to higher recognition rates.

7. REFERENCES

[1] Biing,H.J. and Sadaoki,F. 2000, Automatic recognition and understanding of Spoken launguage- A first step towards natural human-machine communication, Proceedings of the IEEE Vol.88.

[2] Rajoriya, D.K., Anand, R.S. and Maheshwari R.P. 2010, Hindi paired word recognition using probabilistic neural network, International Journal Computational Intelligence studies, Vol.1.

[3] Cowie, R. and Cornelius, R.R. 2003, Describing the emotional states that are expressed in speech, Speech Communication, Elsevier, Vol. 40.

[4] Rajoriya, D.K., Anand, R.S. and Maheshwari, R.P. 2011,Spoken Paired Word Pattern Classification Using Whole Word Template, TECHNIA- International Journal of Computing Science and Communication Technologies, Vol.3.

[5] Saon, G. and Padmanabham, M. 2001,Data-driven approach to designing compound words for continuous speech recognition, IEEE Transactions on Speech and Audio Processing,Vol. 9,No.4, pp.327-332.

[6] Sasaki, S. and Matsumoto, I. 1994, Voice activity detection and transmission error control for digital cordless telephone system, IEICE Trans.Commun., Vol. E77B, no. 7,1994; pp. 948–955.

[7] Hariharan, R., Hakkinan, J. and Laurila, K.200), Robust end of utterance detection for real time speech recognition applications, IEEE International conference on Acoustics, Speech and Signal Processing.

[8] Singh, A., Rajoriya, D.K. and Singh, V. 2012, Database Development and Analysis of Spoken Hindi Hybrid Words Using Endpoint Detection, International Journal j of Electronics and Computer Science Engineering, Vol.1.

[9] Bullington, K. and Fraser, J.M. 1959, Engineering aspects of TASI, Bell Syst.Tech. J'., pp. 353–364.

[10] Ying, G.S.,Mitchell, C.D.,Jamieson, L.H. 1992, Endpoint Detection of Isolated Utterances Based on A Modified Teager Energy Measurement. In Proc. IEEE ICASSP-92, pp.732-pp.735.

.

[11] Atal, B.and Rabiner, L.R. 1976,A pattern recognition approach to voiced-unvoiced-silence classification with applications to speech recognition, Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions on , Volume: 24 , Issue: 3 , Pages: 201 – 212

[12] Childers, D.G.,Hand, M. and Larar J.M. 1989, Silent and Voiced/Unvoied/ Mixed Excitation(Four-Way), Classification of Speech, IEEE Transaction on ASSP, Vol-37, No-11, pp. 1771-74.

[13] Halverson, D.R. 1995 , Robust estimation and signal detection with dependent nonstationary data, Circuits Syst. Signal Process., vol. 14, no.4,pp.465-472.

[14] Rabiner, L.R. and Juang, B.H 1993, Fundamentals of speech recognition, 1st Indian Reprint, Pearson Education.

[15] Wilpon, J.G.,Rabiner, L.R. and Martin, T. 1984, An improved word-detection algorithm for telephone-quality speech incorporating both syntactic and semantic constraints, AT&T Bell Labs. Tech. J., vol. 63, pp. 479–498.

[16] Chengalvarayan, R. 1999, Robust energy normalization using speech/non speech discriminator for German connected digit recognition, in Proc.Eurospeech'99, Budapest, Hungary, pp. 61–64.

[17] Rabiner, L.R. and Sambur, M.R. 1975, An algorithm for determining the endpoints of isolated utterance's, Bell Syst. Tech. ***J.***, vol. 54, pp. 297–315.

[18] Junqua, J.C.,Reaves, B. And Mak B. 1991, A study of endpoint detection algorithms in adverse conditions: Incidence on a DTW and HMM recognize ,in Proc. Eurospeech, pp. 1371–1374

[19] Qi Li, Zheng, J., Tsai, A.and Zhou, Q. 2002, Robust Endpoint Detection and Energy Normalization for Real-Time Speech and Speaker Recognition, IEEE Transactions on speech and audio processing, Vol.10, NO.3.

[20] Tanyer, S.G. and Özer, H. 2000, Voice Activity Detection in Non Stationary Noise, IEEE Transactions on speech and audio processing, Vol. 8, NO. 4.

[21] Qi.Li and Tsai, A. 1999, A language- independent personal voice controller with embedded speaker verification, in Proc. Eurospeech'99, Budapest, Hungary.