Volume 4, Issue 2, February 2014 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Voice Recognized Word Counter

Ripal Patel, Bhargav Goradiya, Anish Vahora, Ankur Dhami, Sandip Gangani
ET Dept. BVM ENGG.College
India

Abstract— The voice recognized word counter is a system that is used to count the particular word. It recognizes the isolated spoken words that we want to count. It then performs the counting operations, and displays the final answer on display. Voice recognition systems have a very strong probability of becoming a necessity in the workplace in the future. Such systems would be able to improve productivity and would be more convenient to use. The idea of a hardware that can recognize any person's voice without the training time involved in currently employed systems is a very promising one, and possibly a marketable one too.

Keyword- Matlab, voice reorganization, noise removable, word extraction, cross correlation, Filtering, thresholding technique

I. INTRODUCTION

One can speak into a microphone and the computer transforms the sound of your words into wav file to be used by your applications available on your computer that how Speech recognition works. The computer may repeat what you just said or it may give you a prompt for what you are expected to say next. This is the central promise of interactive speech recognition. Early speech recognition programs made you speak in staccato fashion, insisting that you leave a gap between every two words. You also had to correct any errors virtually as soon as they happened, which means that you had to concentrate so hard on the software that you often forgot what you were trying to say.

The new voice recognition systems are certainly much easier to use. You can speak at a normal pace without leaving distinct pauses between words. However, you cannot really use "*natural speech*" as claimed by the manufacturers. You must speak clearly, as you do when you leave someone a telephone message. Remember, the computer is relying solely on your spoken words. It cannot interpret your tone or inflection, and it cannot interpret your gestures and facial expressions, which are part of everyday human communication. Some of the systems also look at whole phrases, not just the individual words you speak. They try to get information from the context of your speech, to help work out the correct interpretation. This is how they can (sometimes) work out what you mean when there are several words that sound similar (such as "to," "too" and "two".)

Here we have recorded 20 signals each of Ram, Sita & Krishna. For training purpose we have used 10 signals and for testing purpose we have used 10 signals. We have defined accuracy for all the tested signals. The block diagram of the overall system is shown in Fig. 1 which comprises of various modules like speech acquisition, filter, word extraction, word counter and voice recognition.

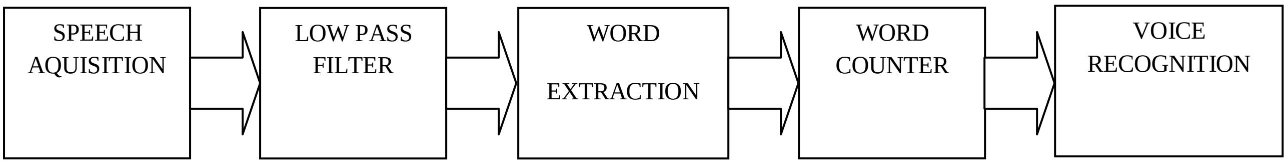


Fig.1 Block Diagram of System

II. RELATED WORK

There are many papers on speech recognition has been published. In [1] Blake S. Wilson and his colleague did better speech recognition with cochlear implants. A cochlear implant system consists of one or more implanted electrodes for direct electrical activation of the auditory nerve, an external speech processor that transforms a microphone input into stimuli for each electrode, and a transcutaneous (rf-link) or per-cutaneous (direct) connection between the processor and the electrodes. High level of speech recognition is achieved with this method. In [2] B. H. Juang & L. R. Rabiner used the hidden markov model (HMM) for the speech recognition. This method is popular because of the inherent statistical framework; the ease and availability of training algorithms for estimating the parameters of the models from finite training sets of speech data; the flexibility of the resulting recognition system in which one can easily change the size, type, or architecture of the models to suit particular words, sounds, and so forth; and the ease of implementation of the overall recognition system. In [3] Chadawan Ittichaichareon and his mates did speech recognition using MFCC. This is currently the accurate method in speech recognition.In [4] Yoshitaka Nishimura and his colleague did speech recognition using multi band spectral features. In most of automatic speech recognition (ASR) system MFCC (mel frequency cepstral coefficients) is used. The problem with using the MFCC is that noise effects spread over all the coefficients even when the noise is limited within a narrow frequency band. If a spectrum feature is directly used, such a problem can be avoided.

III. Proposed Algorithms

A. Speech Actuation

Speech is acquired from a microphone and brought into the development environment for offline analysis. For testing, speech is continuously streamed into the environment for online processing. During the training stage, it is necessary to record repeated utterances of each digit in the dictionary. For example, we repeat. The word ‘one’ many times with a pause between each utterance. Using *wavrecord* and *wavwrite* functions in MATLAB with a standard PC sound card, we capture six seconds of speech from a microphone input at 8192 samples per second. We save the data to disk as ‘ram.wav’as shown in Fig.2

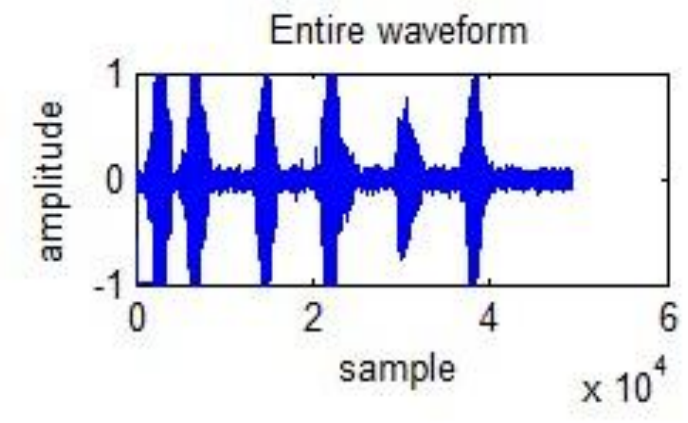


Fig.2 Speech Signal

B. Noise Remove

In applications that use filters to shape the frequency spectrum of a signal such as in communications or control systems, the shape or width of the roll-off also called the "transition band", for a simple first-order filter may be too long or wide and so active filters designed with more than one "order" are required. These types of filters are commonly known as "High-order" or "nth-order" filters.

The complexity or filter type is defined by the filters "order", and which is dependent upon the number of reactive components such as capacitors or inductors within its design. Then, for a filter that has an nth number order, it will have a subsequent roll-off rate of 20n dB/decade or 6N dB/octave.

So a first-order filter has a roll-off rate of 20dB/decade (6dB/octave), a second-order filter has a roll-off rate of 40dB/decade (12dB/octave), and a fourth-order filter has a roll-off rate of 80dB/decade (24dB/octave). As shown in Fig.3, we have applied the Butterworth Low pass filter for noise removal.

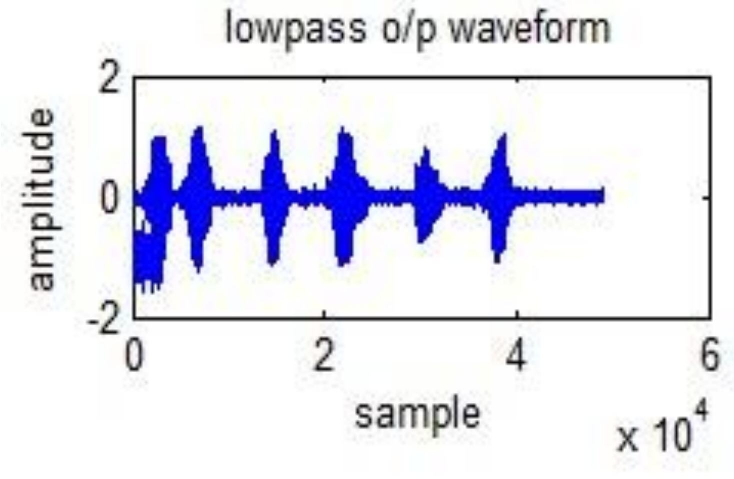


Fig.3 Speech Signal after Low pass Filtering

C. Word Extraction

Now for the recognition of the signal we have to extract the first word from the audio signal. For extracting signal we have to first remove the some blind space in the signal that is always present in the signal to find starting of the word. After removing the blind space we have to extract the signal. For this we have use a run length algorithm mechanism to find the end of first word then we will extract that part of signal for recognition purpose. As shown in Fig.4, We have extracted the first word from the input speech signal. We have extracted words Ram. Sita, Krishna as shown in fig.5 to 7.

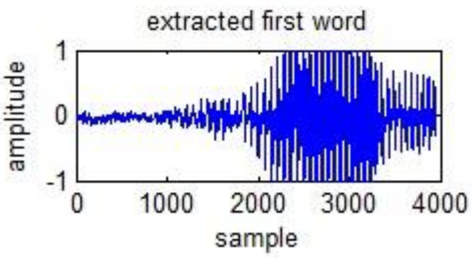


Fig.4 Extracted First Word

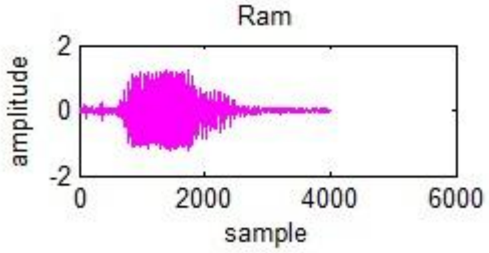


Fig.5 Extracted 'Ram' Word

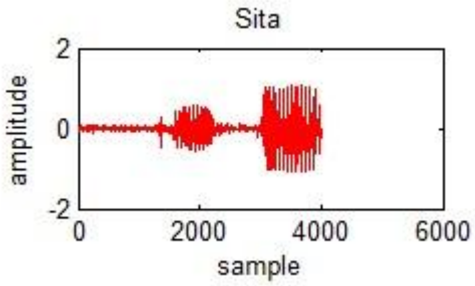


Fig.6 Extracted 'Sita' Word

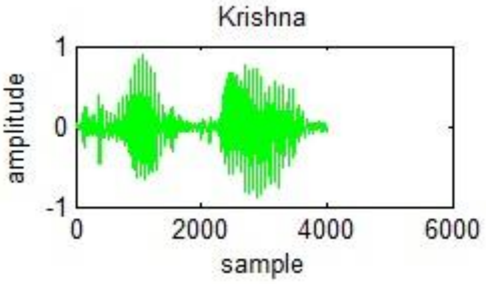


Fig.7 Extracted 'Krishna' Word

D. Thresholding Process

To Count the words spoken we have to count the peaks in the recorded audio signals. For that thresholding technique is used in that we have to set the values of peak of the signals. If the speech signal is greater than that peak value than it is given value '1' otherwise '0'.We have applied threshold level of 2/3rd of the peak value of the signal as shown in fig 8.

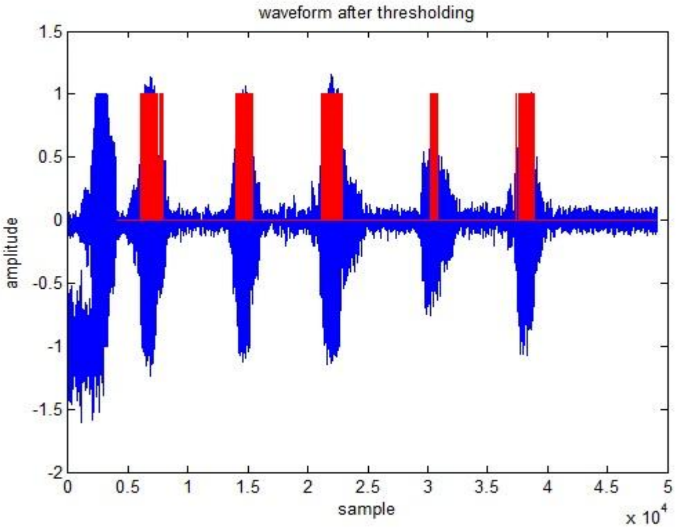


Fig.8 Signal after Thresholding Process

E. Word Count and Crosscorrelation

For counting words we have count the space between two words using Run Length Algorithm by counting the maximum number of zeros. If number of zeros is greater than 4000 then we will increment the word count. Finally decrementing the total word count one by one we will get the word count of the speech signal. In Matlab cross-correlations are computed with the function xcorr which works in the Frequency domain.

The Crosscorrelator returns the cross-correlation sequence for two discrete-time deterministic inputs. This object can also return the cross-correlation sequence estimate for two discrete-time, jointly wide-sense stationary (WSS), random processes. We often use correlation to search for similar signals that are repeated in a time series – this is known as matched filtering. Because the correlation of two high amplitude signals will tend to give big numbers, one cannot determine the similarity of two signals just by Comparing the amplitude of their cross correlation. Therefore we have to use normalized cross correlation. Here we have used the cross correlation to correlate the words spoken by the user and see which word is most correlate with the spoken words and that word will be counted.

The following table 1 shows the value of correlation between different signals and we can see that the same signals is highest compare to other different signal. These signals are shown from Fig.9 to Fig.12.

TABLE I
CORRELATION OF DIFFERENT VOWELS

Signal 1	Signal 2	correlation
m01ah.wav	m01ah.wav	313.4834
m01ah.wav	m01ae.wav	34.5304
m01ah.wav	m01aw.wav	9.7971
m01ah.wav	m01eh.wav	18.1333

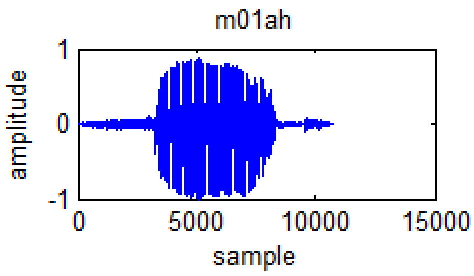


Fig.9 m01ah.wav file

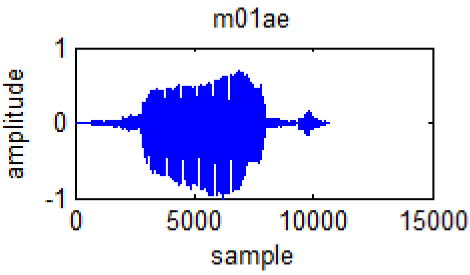


Fig.10 m01ae.wav file

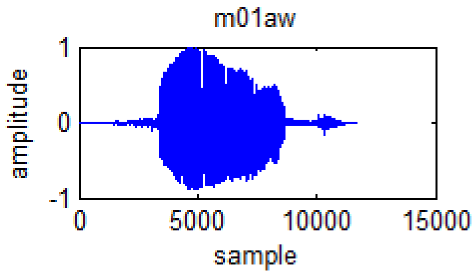


Fig.11 m01aw.wav file

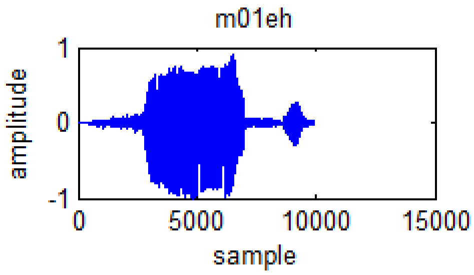


Fig.12 m01eh.wav file

IV. RESULTS

After running the algorithm, the results are obtained. The system requires the user to record words which are stored in the database. After that the system saves the recorded voice into a .wav file. The input word is then recognized as the word corresponding to the extracted words as shown in fig.13. Then by applying thresholding and Run Length algorithm we can count the words of recorded speech signal as shown in Fig.14. The accuracy of the system is defined in the below table 2.

TABLE II
ACCURACY OF PROPOSED ALGORITHM

Word	Train Word	Test Word	True Accurate	Accuracy (%)
Ram	20	10	8	80
Sita	20	10	7	70
Krishna	20	10	7	70
Total	60	30	22	73.33

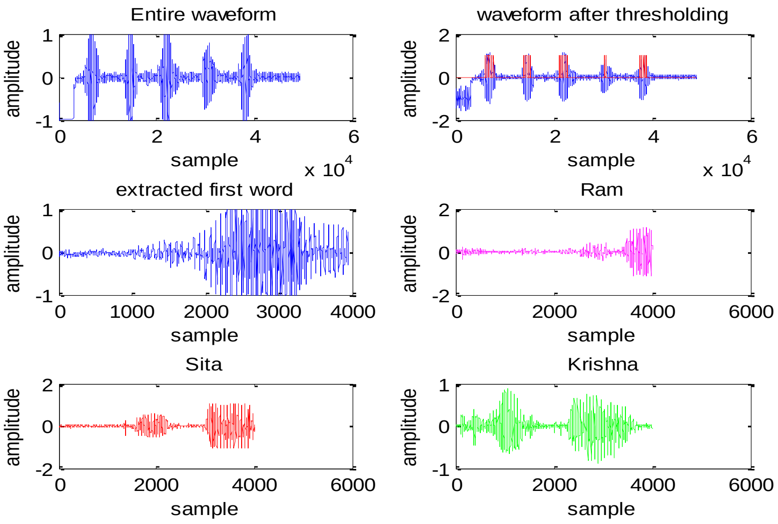
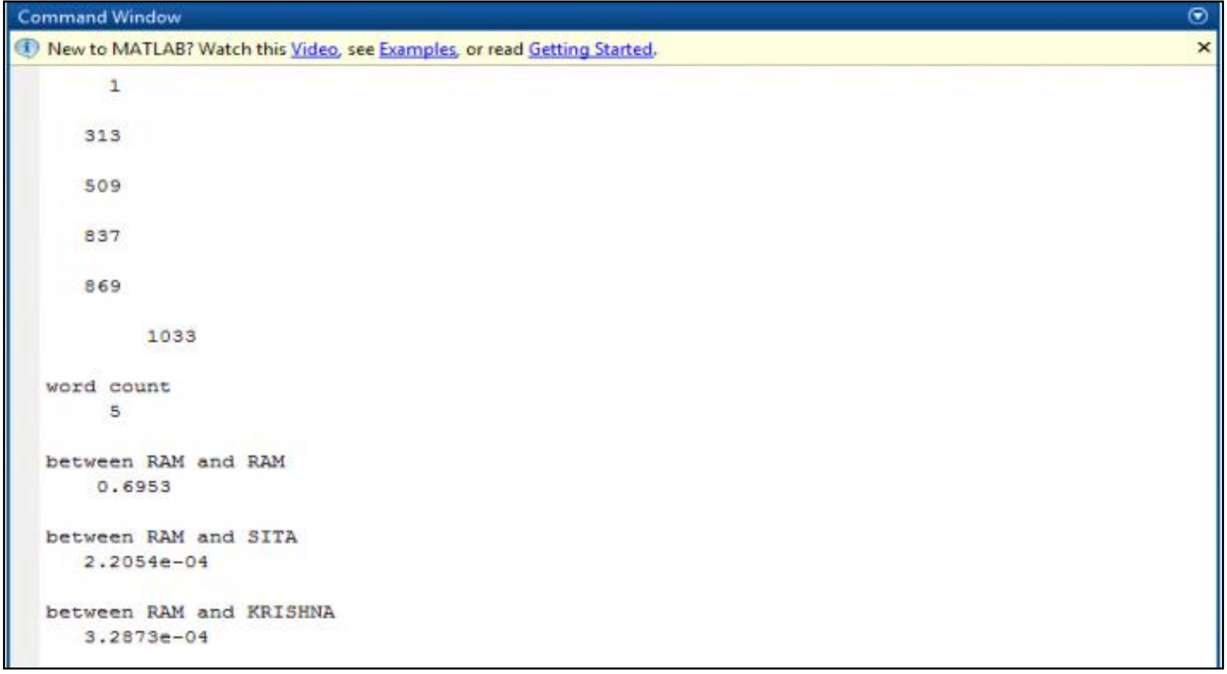


Fig.13 Final resulted waveforms


Command Window
New to MATLAB? Watch this Video, see Examples, or read Getting Started.

```
     1

   313

   509

   837

   869

        1033

word count
     5

between RAM and RAM
    0.6953

between RAM and SITA
   2.2054e-04

between RAM and KRISHNA
   3.2873e-04
```

Fig.14 Result of word count and cross correlation

V. Conclusion

This paper has shown speech recognition and a word count algorithm for isolated words. The results showed a promising speech recognition and word count module. Meanwhile, Thresholding and Run Length Algorithm are used to count the words spoken by the user. Recognition with about 73.33% accuracy can be achieved using this method, which can be further increased with further research and development.

REFERENCES

[1] Better speech recognition with cochlear implants, Blake S. Wilson, Charles C. Finley, Dewey T. Lawson, Robert D. Wolford, Donald K. Eddington & William M. Rabinowitz , *Nature* 352, 236 - 238 (18 July 1991); doi:10.1038/352236a0

[2] Hidden Markov Models for Speech Recognition, B. H. Juang & L. R. Rabiner, pages 251-272, Speech Research Department, AT&T Bell Laboratories , Murray Hill , NJ , 07974

[3] Speech Recognition using MFCC ,Chadawan Ittichaichareon, Siwat Suksri and Thaweesak Yingthawornsuk, International Conference on Computer Graphics, Simulation and Modeling (ICGSM'2012) July 28-29, 2012 Pattaya (Thailand)

[4] Noise-robust speech recognition using multi-band spectral features Yoshitaka Nishimura, Takahiro Shinozaki, Koji Iwano and Sadaoki Furui, J. Acoust. Soc. Am. 116, 2480 (2004)