Word Recognition Using Barthannwin Wave Filter and Neural Network

Abhilasha Singh Rathor and Pawan Kumar Mishra

Abstract Speech Recognition is the process of automatically recognizing the spoken words of person based on information in speech signal. Recognition technique makes it possible to the speaker's voice to be used in verifying their identity and control access to services such as voice dialing, telephone shopping, banking by telephone, database access services, information service, voice mail, remote access to computers and security control for the confidential information areas. Digital filter plays an important role in digital signal processing applications. Digital filter can also be applied in speech processing applications, such as speech enhancement, speech filtering, noise reduction and automatic speech recognition among others. Filtering is a widely researched topic in the present era of communications. As the received signal is continuously corrupted by noise where both the received signal and noise change continuously, and then arises the need for filtering. This paper provides introduction to Barthannwin Wave Filter based on Barthannwin window for speech signal modeling suited to high performance and robust isolated word recognition. It provides efficient performances with less computational complexity. A Barthannwin Wave filter is designed based on the estimated noise statistics, and it is useful for noise reduction of the speech. The proposed filtering scheme outperforms other existing speech enhancement methods in terms of accuracy in a word recognition system.

Keywords Speech recognition · Speech processing · Word recognition · Filter design

A. S. Rathor (✉) · P. K. Mishra
Faculty of Technology, UTU, Dehradun, India
e-mail: abhilasha.rathor@gmail.com

P. K. Mishra
e-mail: pawantechno@gmail.com

1 Introduction

Speech is considered to be a primary mode of communication among human beings. It is also a most efficient and natural form of exchanging information among human [1]. Speech comes so naturally to everybody that we don't realize that speech is very difficult phenomenon to understand. We as human being understand speech very easily but computers have trouble to understand it. Computers are very good at recognizing and understanding instructions given to them but they have trouble in understanding speech. There are several problems in recognizing speech such as: (1) None of the words sound the same. (2) All speakers say the word in different way. (3) Context of language is meaningless.

Speech recognition systems can be categorized in different classes on the basis of what type of word they can recognize [1]. Isolated word can be defined as single word at a time. These recognizers generally involve every utterance to include quiet on both sides of sample windows [1]. Connected word can be said as group of words spoken together with minimum pause in between [1]. Continuous speech is similar to speaking a line or a paragraph [1]. Spontaneous speech is when human being speaks in natural way with or without pause between the words [1]. Digital representation is required to process a Signal, thus speech processing is regarded as a special area of digital signal processing (DSP), which can be applied to speech signal. Speech processing contains the acquirement, manipulating, storing, transferring and obtaining output of speech signals. Since the 1960s computer scientists have been researching ways and means to make computers able to record, interpret and understand human speech [1]. Speech Recognition, SR is conversion of spoken word into text. Some speech recognition systems use training" and some do not. In trained SR systems every individual speaker reads a part text into the speech recognition system. After that these systems evaluate the specific voice of person after that they use this voice to fine tune the person's speech recognition, which results in more perfect transcript. Speaker Dependent systems are those systems which are required to be trained while Speaker Independent systems are those systems which are not required to be trained. Speech Recognition can be said as special case of pattern recognition. Training and Testing are two phase in supervised pattern recognition. In the training phase, the parameters of the arrangement model are anticipated using a large number of class examples (Training Data) while the testing or recognition phase, the attributes of test pattern is matched with the trained model of each and every class. The test pattern is said to belong to that whose model matches the test pattern best [1]. Speech Recognition systems performance is usually calculated in terms of speed and accuracy. Speed is measured along with real time factor while Accuracy is considered as Word error rate (WER). Additional measurements of accuracy are Command Success Rate (CSR) and Single Word Error Rate (SWER) [2]. Idea of speech to text can be hard to implement for intellectually disabled persons because of the fact that there are rare chances that anyone will try to learn these technologies to teach the persons with the disabilities [3].

1.1 Problem Definition

When a word is spoken, it is assumed that speech segments can be unfailingly separated from the non-speech segments. The process of separating the speech signals of an utterance from the background noise, i.e., the non-speech segments obtain while recording the signal, is known as endpoint detection. Isolated word recognition systems are used to accurately detect the endpoints of spoken words. It is important for two reasons: (1) Reliable word recognition is importantly dependent on the accurate endpoint detection. (2) When the endpoints are accurately located the computation for processing the speech becomes minimum. There are several problems in accurately locating the endpoints of isolated words while recordings. The problems in endpoint detection occur from transients connected with the speaker or the transmission system. Background noises also complicate the endpoint detection problem significantly. Thus accurate endpoint detection method is very essential component of word recognition. The necessary components of a speech recognition system are feature extraction, pattern comparison, and decision rule. While Endpoint detection is performed in the processing. Noise-robust speech endpoint detection is significant to get practical speech recognition in a noisy real-world environment [4]. The endpoints can be founded in explicitly, implicitly, or hybrid manner.

1.2 Literature Survey

In past years a number of dissimilar methodologies have been proposed for continuous speech and isolated word recognition. These are usually grouped in two classes called as speaker-dependent and speaker-independent. Speaker dependent methods involve training a system to recognize each vocabulary words uttered single or multiple times by specific set of speakers while for speaker independent systems such training methods are not applicable and words are recognized by analyzing their intrinsic acoustical properties. Recent research is focusing on three main features [5]: Large vocabulary size, Continuous Speech Capabilities, Speaker Independent System. Many systems use Hidden Markov models (HMMs) widely. While other uses Neural Networks, some also use Dynamic Time Wrapping and many more techniques [6, 7].

Al-Alaoui et al. [8] analyzed and discussed the applicability of artificial neural networks to speech recognition. A total number of 200 vowel signals from individuals with different gender and races were recorded. The filtering process was performed using the wavelet approach to de-noise and to compress the speech signals. Kotnik et al. [9] proposed a multiconditional robust mel frequency Cepstral coefficients feature extraction algorithm as an alternative to commonly used symmetrical Hamming window. He also proposed Cosine window (hHCw) in the preprocessing stage. Betkowska et al. [10] discussed the problem of speech

recognition in the presence of nonstationary sudden noise, like in home environments. The proposed FHMMs achieved better recognition accuracy than clean-speech HMMs for different SNRs. The overall relative error reduction given by phoneme FHMMs was 12.8 % compared to that given by clean-speech HMMs. Lim et al. [11] implemented a new pattern classification method, by Neural Networks trained using the Al-Alaoui Algorithm. The new method gave comparable results to already implement HMM method for recognition of words, and it has overcome HMM in the recognition of sentences. He compared two different methods for automatic Arabic speech recognition for isolated words and sentences. The KNN classifier gave better results than the NN in the prediction of sentences. Muda et al. [12] discussed two voice recognition algorithms Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW), which are important in improving voice recognition performance. Pandey et al. [13] discussed how binary recurring neural network can be used to solve the problem of recognizing sentences having similar meaning but different lexico-grammatical structures in English. Abushariah et al. [14] designed and implemented English digits speech recognition system using Matlab (GUI). It was based on Hidden Markov Model (HMM), which provided extremely trustworthy technique for speech recognition. Rafieee and Khazaei [15] proposed a new model for a noise robust Automatic Speech Recognition (ASR) based on parallel branch Hidden Markov Model (HMM) structure with novel approach for robust speech recognition. The characteristics of a novel model are presented by exploring vibrocervigraphic and ectromyographic ASR methods and some other successful approaches to achieve the best results. Paul et al. [16] proposed a methodology for automated recognition of isolated words independent of speakers. She computed ZCR by partitioning audio signal into segments and calculates the number of times the signal crosses zero amplitude level within each segment. Wijoyo [17] implemented speech recognition system on mobile robot for controlling movement of the robot. They used Linear Predictive Coding (LPC) and Artificial Neural Network (ANN) for speech recognition system. LPC method is used for extracting feature of a voice signal and ANN is used as the recognition method. Back propagation method is used to train ANN. Experimental results show that highest recognition rate that can be achieved by this system is 91.4 %. Ittichaichareon et al. [18] discussed an approach of speech recognition by using the Mel-Scale Frequency Cepstral Coefficients (MFCC) extracted from speech signal of spoken words. Based on experimental database of total 40 times of speaking words collected under acoustically controlled room, MFCC extracts have shown the improvement in recognition rates significantly when training the SVM with more MFCC samples by randomly selected from database, compared with ML.
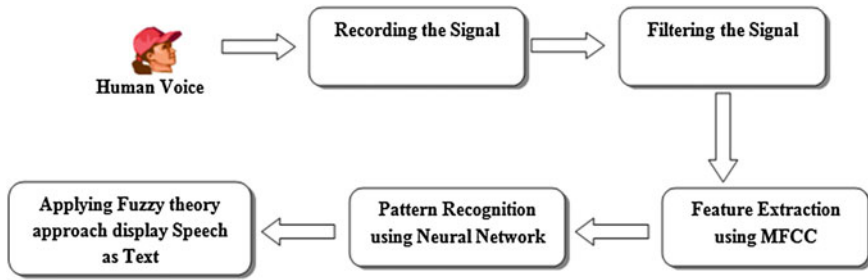
Fig. 1 Word recognition system design

2 Word Recognition System Design

The basic idea is to design a filter so that we can remove the noises and unwanted signal from the recorded voice and obtain high efficiency in Word Recognition. The scheme used for Word Recognition along with the algorithm of proposed Barthannwin Wave Filter is discussed. Word Recognition is an area of Speech Recognition, where we recognize the words spoken by several speakers with the words stored in our database. Word recognition requires the extraction of features from the recorded utterances followed by a training phase [19, 20]. There are several steps in word recognition system, our main focus in on filtering the signal and designing an efficient filter for word recognition system. A word recognition system design is shown in Fig. 1.

For recording the signal we have used MATLAB frontend. After recording the signal for filtering process Barthannwin Wave Filter is implemented. Filtering is a process that removes some unwanted distortion or noise from a signal. It aims at removing some frequencies in order to suppress interfering signals and reduce background noise. Filters are widely used in signal processing and communication systems in applications such as channel equalization, noise reduction, radar, audio processing, video processing, biomedical signal processing, and analysis of economic and financial data.

The primary functions of filters are: (1) To confine a signal into a prescribed frequency band as in low-pass, high-pass, and band-pass filters. (2) To decompose signal into two or more sub-bands as in filter-banks, graphic equalizers, sub-band coders, frequency multiplexers. (3) To modify frequency spectrum of a signal as in telephone channel equalization and audio graphic equalizers. (4) To model input–output relationship of a system such as telecommunication channels, human vocal tract, and music synthesizers.

Filter designing means to select the filter coefficients in such a way so that the system has specific characteristics. These characteristics are stated as filter specifications. Mostly the time filter specifications are known as frequency response of the filter. We also need to choose the transfer function and filter structure. Mapping the transfer function to the filter structure gives the element values of analog filters

elements and the coefficients of digital filter. The methods to find the coefficients of a filter from its frequency specifications are: (1) Window design method (2) Frequency Sampling method (3) Weighted least squares design (4) Parks-McClellan method (5) Equiripple FIR filters design using the FFT algorithms. We have used Window Design Method since in filter design, windows are typically used to reduce unwanted ripples of the filter in the frequency response. FFT windows reduce the effects of leakage but cannot eliminate leakage entirely. In effect, they only change the shape of the leakage. In addition, each type of window affects the spectrum in a slightly different way. Many different windows have been proposed over time, each with its own advantage and disadvantage relative to the others. Some are more effective for specific types of signal types such as random or sinusoidal. Some improve the frequency resolution, that is, they make it easier to detect the exact frequency of a peak in the spectrum. Some improve the amplitude accuracy, that is, they most accurately indicate the level of the peak. The best type of window should be chosen for each specific application. We have chosen Barthannwin window which is a modified form of Bartlett-Hann window.

2.1 Proposed Filter-Barthannwin Wave Filter

Barthannwin Wave Filter is a high pass FIR filter which is designed by using Barthannwin window. We have designed a FIR filter because of its following basic characteristics:

1. Linear phase characteristic
2. High filter order (more complex circuits); and
3. Stability

Barthannwin window is a combination of two windows Bartlett Window and Hann (also known as Hann) Window.

Bartlett Window: The Bartlett window is same as a triangular window. Bartlett window ends with 0's at samples 1 and n, while at those points the triangular window is nonzero. The center L-2 points of Bartlett (L) are equivalent to triang (L-2) for L odd. If we specify L = 1 for a one-point window, the value 1 is returned.

`w = bartlett(L)` returns an L-point Bartlett window in the column vector `w`, where `L` should only be a positive integer. The Bartlett window coefficients can be computed as follows:

$$w(n) = \begin{cases} \frac{2n}{N}, & 0 \leq n \leq \frac{N}{2} \\ 2 - \frac{2n}{N}, & \frac{N}{2} \leq n \leq N \end{cases}$$

The length of window $L = N+1$.

Hann Window: The Hann window has the shape of one cycle of a cosine wave with 1 added to it so it is always positive. After that the sampled signal values are

multiplied by the Hann function. The ends of the time record are forced to zero regardless of what the input signal is doing. The Hann window should always be used with continuous signals, but must never be used with transients. The reason is that the window shape will distort the shape of the transient, and the frequency and phase content of a transient is intimately connected with its shape.

`w = hann(L)` returns an L-point symmetric Hann window in the column vector w. `L` should be a positive integer. Hann window coefficients are computed from the equation.

$$w(n) = 0.5\left(1 - \cos\left(2\pi\frac{n}{N}\right)\right), 0 \le n \le N$$

The window length is $L = N + 1$.

Barthannwin Window: This window consist main lobe at its origin and decaying side lobes on both sides asymptotically. It is a result of linear combination of Hann and weighted Bartlett windows with near side lobes lower than both Bartlett and Hann and with far side lobes lower than both Bartlett and Hamming windows. The main lobe width of the modified Bartlett-Hann window is not increased relative to either Bartlett or Hann window mainlobes.

`w = barthannwin(L)` returns an L-point modified Bartlett-Hann window in the column vector w. The following equation is used to compute the coefficients of a Modified Bartlett-Hann window is

$$w(n) = 0.62 - 0.48\left|\left(\frac{n}{N} - 0.5\right)\right| + 0.38\left(2\pi\left(\frac{n}{N} - 0.5\right)\right)$$

where $0 \le n \le N$ and the window length is $L = N + 1$.

After deciding which window to be used a high pass FIR filter was designed using Barthannwin Window known as Barthannwin Wave Filter.

For the feature extraction of the speech signal the Mel-frequency cepstrum coefficient is used in this model. In speech processing, mel-frequency cepstrum (MFC) represent the short-term power spectrum of the sound signal, which is based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency. The feature measurements of speech signals are typically extracted using one of the following spectral analysis techniques: MFCC Mel frequency filter bank analyzer, LPC analysis or discrete Fourier transform analysis. Currently the most popular features are Mel frequency Cepstral coefficients MFCC. There are two ways to analyze the MFCCs: (a) as a filter-bank processing adapted to speech specificities and (b) as a modification of the conventional cepstrum, a well-known deconvolution technique based on homomorphic processing [33]. The block diagram for calculating MFCCs is shown in Fig. 2.

Neural networks are composed of simple elements operating in parallel, inspired by biological nervous systems. The network function is determined largely by the connections between elements. We can train a neural network to perform a particular function by adjusting the values of the connections (weights) between elements. Neural networks are adjusted, or trained, so that a particular
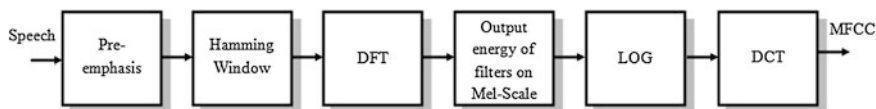
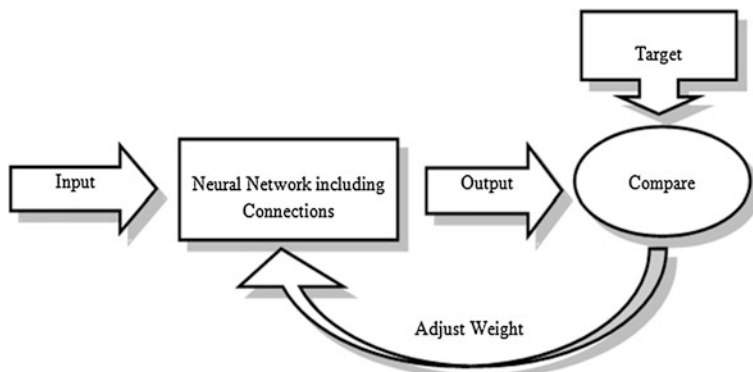**Fig. 2** MFCC calculation

Fig. 3 Neural network training

input leads to a specific target output. There, the network is adjusted, based on a comparison of the output and the target, until the network output matches the target.

Neural networks have been trained to perform complex functions in various fields of application including pattern recognition, identification, classification, speech, vision and control systems. The supervised training methods are commonly used, but other networks can be obtained from unsupervised training techniques or from direct design methods. Radial basis functions are dominant techniques used for interpolation in the multidimensional space. A RBF is a function which has constructed into a distance criterion with respect to a center. RBF networks have two layers of processing: In the first, input is mapped onto each RBF in the 'hidden' layer. The RBF chosen is usually a Gaussian.

Radial basis function (RBF) networks typically have three layers: an input layer, a hidden layer with a non-linear RBF activation function and a linear output layer [39]. Architecture of a radial basis function network is shown in Fig 3. An input vector x is used as input to all radial basis functions, each with different parameters. The network output is a linear combination of outputs from radial basis functions.

A common method of training a neural net in which the initial system output is compared to the desired output and the system is adjusted until the difference between the two is minimized. It is a controlled learning method, and simplification of the delta rule. For making the training set it requires a dataset of desired

output for different inputs. Back propagation needs that the activation function which is used by artificial neurons be differentiable. Back propagation networks are essentially multilayer Perceptrons (typically with one input, hidden, and output layer). The back propagation learning algorithm can be divided into two stages: propagation and weight update.

2.2 Steps for Word Recognition

Steps for Word Recognition process are explained below.

Step 1: Begin
Step 2: Get The Input Data   */ Record the Samples
*/ Check if Samples are correct or not, if samples are not correct, record again
Step 3: Check The Input Data
*/ Apply designed filter so that important information is retained and rest is discarded
Step 4: Apply Barthannwin Wave Filter
*/ Extract the features of the signal using MFCC
Step 5: Extract Features
*/ Create the network where P denotes the Input and T denotes the Target
Step 6: Create The Network
        P ← Input
        T ← Target
*/ Train the Neural Network
Step 7: Training The Neural Network
*/ Apply the Fuzzy Set-Theoric Approach to match Input to Target
Step 8: Apply Fuzzy Set-Theoric Approach
*/ If match found display the word
Step 9: If Output Is Equal To Target
        Display Word
        Play Voice
        End if
*/ If word not found Display that word is not found
Step 10: If Word Not Found Display "Word Not Found"
Step 11: End

2.3 Algorithm for Barthannwin Window

1. Begin
2. if number of input arguments < 1
        Use barthannwin function

End if
3. if length of Signal < 0
        Display "Error"
   End if
4. Compute L ← round(L)   */L denotes the length of signal
5. Compute N ← L − 1   */N denotes maximum Order
6. For n ← 0 to N
        Compute w = 0.62 – 0.48|(n/N – 0.5)| + 0.38 cos[2Π(n/N − 0.5)]
   End
7. Compute w ← w'
8. End

2.4 Algorithm for Barthannwin Wave Filter

1. Begin
   */All frequency values are in Hz.
2. Set Fs ← 32000   */Fs denotes sampling frequency
3. Set N ← 32000   */N denotes Order
4. Set Fc ← 10800   */Fc denotes Cutoff frequency
   */Sampling Flag, 'scale' to normalize the filter so that the magnitude response of
   the filter at the center frequency of the passband is 0 dB. */
5. Set Flag ← 'Scale'
   */Create the window vector for the design algorithm.
6. Set win ← barthannwin (N + 1)
   */calculate the coefficients using the FIR1 function.
7. Compute b = fir1(N, Fc/(Fs/2), 'high', win, flag);
   */'high' is for a highpass filter with cutoff frequency Fc/(Fs/2).
   */returns a discrete-time, direct-form finite impulse response (FIR) filter, Hd,
   with numerator coefficients, b. */
8. Compute Hd = dfilt.dffir(b);
9. End

3 Experimental Results

The experiments are executed in order to assess performance of the newly developed Barthannwin Wave filter applied on fuzzy word recognition. In interpreting the results, consideration must be taken since the behaviour of the algorithm is highly dependent on the parameter values. The results are only meaningful for the specific test signal set. For other test signals, the performance may be different.
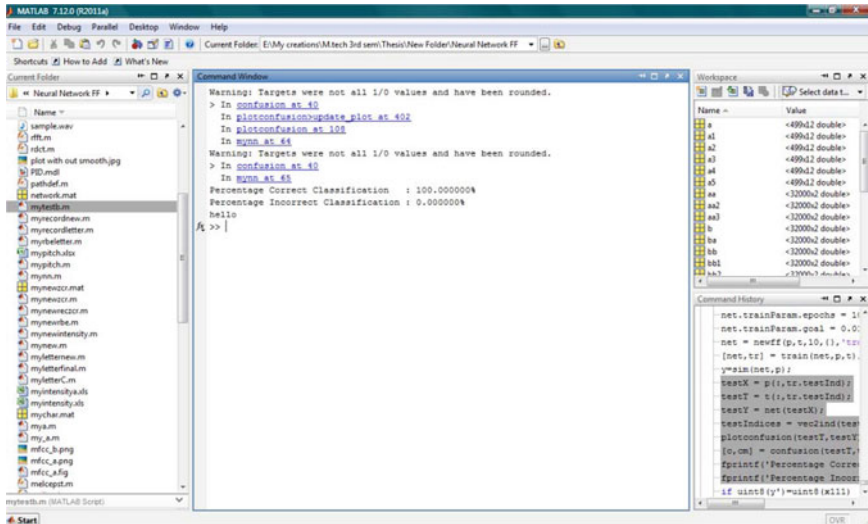
Fig. 4 Output window for word "Hello" with 100 % correct classification

The new filter was implemented in a word recognization system and its accuracy is shown with the help of different words spoken by different age group.

The results obtained are shown in Fig. 4.

The 5 words were recorded that are "Hello", "How", "Are", "You" and "Fine" using microphones. 2 s time was allotted to record a word. System was trained for these 5 words with 1 male and 2 female voice. After that different samples were collected from people belonging to different age groups and accuracy of system was calculated.

All the results of word "Hello" in both Male and Female voices is as shown. Which shows more than 97 % of accuracy is achieved in all samples of Male voice and 90 % of accuracy is achieved in the samples of Female voice (Tables 1, 2).

4 Conclusion

The basic of Barthannwin Window, system identification and Barthannwin Wave Filter using Barthannwin Window was designed. A new word recognition algorithm for speech corrupted by slowly varying additive background noise based on Barthannwin Wave filtering was developed. It was demonstrated that this new algorithm can enhance the accuracy of word recognition system. The algorithm was implemented and evaluated. The parallel version proves that speech enhancement algorithms can profit from parallel computing techniques. Nevertheless, the rather modest quality improvement of the resulting speech over alternative methods would not by itself justify the significant increase in

Table 1 Accuracy of word "hello" in male voice

Number of samples Age group	10–15 years	15–25 years
10	99 %	98.6 %
50	98 %	98.2 %
80	97.6 %	97.2 %

Table 2 Accuracy of word "hello" in female voice

Number of samples Age group	10–15 years	15–25 years
10	95 %	95.6 %
50	93 %	93.4 %
80	92.6 %	92.2 %

computational complexity. This work also demonstrates the use of a model based approach for word recognition. In the case studied, to identify the model parameters correctly and reliably was found to be the critical part of the algorithm. Finally, it is important to keep in mind that for speech enhancement systems designed for human listener, it is the human listener who is the ultimate judge. The conversion to an optimization problem is a design and engineering decision and must be always verified by listening tests. As far as enhancement systems for speech recognition are concerned, their design criterion should closely match the recognizer structure and therefore, it seems unlikely that a single enhancement algorithm would perform well for both tasks. Constant progress is being done in the domain of speech enhancement but it is still a challenging and rewarding field where full of problems are waiting to be solved.

References

1. Gaikwad, S.K., Gawali, B.W., Yannawar, P.: A review on speech recognition technique. Int. J. Comput. Appl. (0975–8887) **10**(3), 16–24 (2010)
2. Singh, G.N.: Literature review on automatic speech recognition. Int. J. Comput. Appl. **41**(8), 42–50 (2012)
3. Hawley, M.S., Green, P., Enderby, P., Cunningham, S., Moore, R.K.: Speech technology for e-inclusion of people with physical disabilities and disordered speech. In: INTERSPEECH 2005, pp 445–448, Sept 2005
4. Kitayama, K., Goto, M., Itou, K., Kobayashi, T.: Speech starter: noise-robust endpoint detection by using filled pauses. In: Proceedings of the 8th European Conference on Speech Communication and Technology (Eurospeech 2003), pp. 1237–1240, Sept 2003
5. Vimala, C., Radha, V.: A review on speech recognition challenges and approaches. World Comput. Sci. Inf. Technol. J. (WCSIT) **2**(1), 1–7 (2012)
6. Das, B.P., Parekh, R.: Recognition of isolated words using features based on LPC, MFCC, ZCR and STE, with neural network classifiers. **2**(3), 854–858 (2012), ISSN: 2249–6645
7. Chunsheng, F.: From Dynamic Time Warping (DTW) to Hidden Markov Model (HMM). Final project report for ECE742 stochastic decision, March 2009

8. Al-Alaoui, M.A., Al-Kanj, L., Azar, J., Yaacoub, E.: Speech recognition using artificial neural networks and hidden markov models. IEEE Multi. Eng. Educ. Mag. **3**(3), 419–423 (2008)
9. Kotnik, B., Vlaj, D., Kacic, Z., Horvat, B.: Robust MFCC feature extraction algorithm using efficient addictive and convolutional noise reduction procedures. In: ICSLP'02 Proceedings, pp. 445–448 (2002)
10. Betkowska, A., Shinoda, K., Furui, S.: Speech recognition using FHMMs robust against nonstationary noise. In: ICASSP 2007, pp. 1029–1032
11. Lim, C.P., Woo, S.C., Loh, A.S., Osman, R.: Speech recognition using artificial neural networks. In: Web information systems engineering, IEEE, Inc. 2000, pp. 77–86
12. Muda, L., Begam, M., Elamvazuthi, I.: Voice recognition algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) techniques. J. Comput. **2**(3), 138–143 (2010)
13. Pandey, B., Ranjan, S., Shukla, A., Tiwari, R.: Sentence recognition using hopfield neural network. IJCSI Int. J. Comput. Sci. Issues 7(4), 12–17 (2010)
14. Abushariah, A.A.M., Gunawan, T.S., Khalifa, O.O., Abushariah, M.A.M.: English digits speech recognition system based on hidden markov models. In: International Conference on Computer and Communication Engineering (ICCCE 2010), 11–13 May 2010, Kuala Lumpur, Malaysia
15. Rafiee, M.S., Khazaei, A.A.: A novel model characteristics for noise-robust automatic speech recognition based on HMM. In: Proceedings of IEEE International Conference on Wireless Communications, Networking and Information Security (WCNIS), 2010, pp. 215–218
16. Paul, D., Parekh, R.: Automated speech recognition of isolated words using neural networks. Int. J. Eng. Sci. Technol. (IJEST) **3**(6), 4993–5000 (2011)
17. Wijoyo, T.S.: Speech recognition using linear predictive coding and artificial neural network for controlling movement of mobile robot. International Conference on Information and Electronics Engineering IPCSIT vol. 6, pp. 179–183 (2011)
18. Ittichaichareon, C., Suksri, S., Yingthawornsuk, T.: Speech recognition using MFCC. In: International Conference on Computer Graphics, Simulation and Modeling (ICGSM'2012). Pattaya (Thailand), 28–29 July 2012
19. Lipeika, A., Lipeikien, J., Telksnys, L.: Development of isolated word speech recognition system. Informatica **13**(1), 37–46 (2002)
20. Ashouri, M.R.: Isolated word recognition using highorder statistics and time -delay neural networks. In: IEEE Signal Processing Workshop on Higher-Order Statistics, IEEE, Inc., 1997