

# Design and Development of Marathi Speech Interface System

Santosh Gaikwad, Bharti Gawali and Suresh Mehrotra

**Abstract** Speech is the most prominent and natural form of communication between humans. It has potential of being an important mode of interaction with computer. Man-machine interface has always been proven to be a challenging area in natural language processing and in speech recognition research. There are growing interests in developing machines that can accept speech as input. Normal person generally communicate with the computer through a mouse or keyboard. It requires training and hard work as well as knowledge about computer, which is a limitation at certain levels. Marathi is used as official language at government of Maharashtra. There is a need for developing systems that enable human-machine interaction in Indian regional languages. The objective of this research is to design and development of the Marathi speech Activated Talking Calculator (MSAC) as an interface system. The MSAC is speaker-dependent speech recognition system that is used to perform basic mathematical operation. It can recognize isolated spoken digit from 0 to 50 and basic operation like addition, subtraction, multiplication, start, stop, equal, and exit. Database is an essential requirement to design the speech recognition system. To reach up to the objectives set, a database having 22,320 sizes of vocabularies is developed. The MSAC system trained and tested using the Mel Frequency Cepstral Coefficients (MFCC), Linear Discriminative Analysis (LDA), Principal Component Analysis (PCA), Linear Predictive Coding (LPC), and Rasta-PLP individually. Training and testing of MSAC system are done with individually Mel Frequency Linear Discriminative Analysis (MFLDA), Mel Frequency Principal Component Analysis (MFPCA), Mel Frequency Discrete

---

S. Gaikwad (✉) · B. Gawali · S. Mehrotra  
System Communication Machine Learning Research Laboratory (SCM-RL),  
Department of Computer Science and Information Technology,  
Dr. Babasaheb Ambedkar Marathwada University, Aurangabad,  
Maharashtra, India  
e-mail: santosh.gaikwadcsit@gmail.com

B. Gawali  
e-mail: bharti\_rokade@yahoo.co.in

S. Mehrotra  
e-mail: mehrotra\_suresh15j@gmail.com

Wavelet Transformation (MFDWT), and Mel Frequency Linear Discrete Wavelet Transformation (MFLDWT) fusion feature extraction techniques. This experiment is proposed and tested the Wavelet Decomposed Cepstral Coefficient (WDCC) with 18, 36, and 54 coefficients approach. The performance of MSAC system is calculated on the basis of accuracy and real-time factor (RTF). From the experimental results, it is observed that the MFCC with 39 coefficients achieved higher accuracy than 13 and 26 variations. The MFLDWT is proven higher accuracy than MFLDA, MFPCA, MFDWT, and Mel Frequency Principal Discrete Wavelet Transformation (MFPDWT). From this research, we recommended that WDCC is robust and dynamic techniques than MFCC, LDA, PCA, and LPC. MSAC interface application is directly beneficial for society people for their day to day activity.

**Keywords** Human-computer interaction · MSAC · MFCC · PCA · WDCC · LPC

## 1 Introduction

Speech is the most natural and efficient form of exchanging information between human. Automatic speech recognition (ASR) is defined as the process of converting a speech signal to a sequence of words by means of an algorithm implemented by a computer program. Speech recognition systems help users who cannot be able to use the traditional Input and Output (I/O) devices [1, 2]. The man-machine interface using speech recognition has helpful ways to for enable the visually impaired and computer laymen to use the updated technologies [3]. There are growing interests in developing machines that can accept speech as input. Given the substantial research efforts in the speech recognition worldwide and the rate at which computer becomes faster and smaller, we can expect more applications of speech recognition. The concept of machines being able to interact with people in natural form is very interesting. It is desirable to have machine interaction in voice mode in one's native language. This is exclusively important in multilingual country such as India, where a majority of the people is not comfortable with speaking, reading, and listening English language [4]. Research in ASR by machine has attracted a great deal of attention over the past six decades. For centuries, researcher has tried to develop machines that can produce and understand speech as humans do so naturally in native language. Some successful applications of speech recognition are virtual reality, multimedia searches, auto-attendants, IVRS, natural language understanding and many more applications [5–7].

This paper is organized as follows: Sect. 2 describes the related work of this experiment. Section 3 describes the speech recognition with feature extraction techniques. Section 4 explains the design of the Marathi speech interface system. Section 5 describes the experimental results and their extensive application for talking calculator and Sect. 6 gives the concluding remark of paper followed by references.

## 2 Related Work

The main goal of speech recognition is to develop techniques and recognition systems for speech input to the machine. Human beings are comfortable speaking directly with computers rather than depending on primitive interfaces such as keyboards and pointing devices. The primitive interfaces like keyboard and pointing devices require a certain amount of skill for effective usage. Use of mouse requires good hand-eye coordination. It is very difficult for visually handicapped person to use the computer. Moreover, the current computer interface assumes a certain level of literacy from the user. It expects the user to have certain level of proficiency in English apart from typing skill. Speech interface helps to resolve these issues [8]. Computers which can recognize speech in native languages enable the common man to make use of those benefits in education technology [9]. The researcher turns toward performance improvement of speech recognition system for significant innovation of real-time applications [10]. Many factors are affecting the speech recognition such as regional, sociolinguistic, or related to the environment. These create a wide range of variations that may not be modeled correctly (speaker, gender, speaking rate, vocal effort, regional accent, speaking style, nonstationary, etc.), especially when resources for system training are infrequent. Performance is affected by speech variability [11]. The majority of technological changes have been directed toward the purpose of increasing robustness of the recognition system [12].

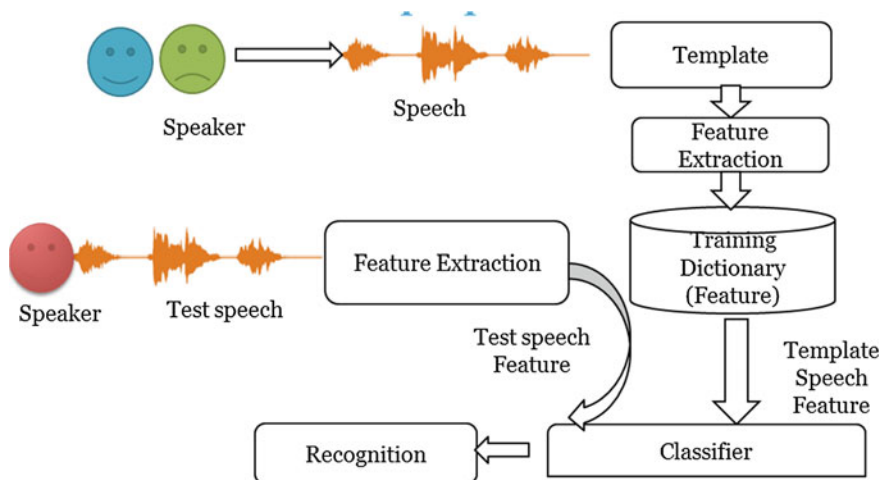
N.S. Nehe et al., proposed an efficient feature extraction method for speech recognition. The features were obtained Linear Predictive Coding (LPC) and Discrete Wavelet Transformation (DWT). The proposed approach provides effective (better recognition rate), efficient (reduced feature vector dimension) features. Continuous Density Hidden Markov Model (CDHMM) has been implemented for system classification. The proposed algorithms were evaluated using an isolated Marathi digits database in the presence of white Gaussian noise [13]. Rajkumar S. Bhosale et al., worked on speech-independent recognition system using multi-class support vector machine and LPC [14].

The research work for Indian languages in speech recognition has not yet grasped to a critical level for a real-time communication as compare to other languages of developed countries. Countable attempts to develop a speech recognition system had been attempted by HP Labs India and IBM research lab, Google, IIT Powai, CDAC Pune [15, 16].

However, there is lots of opportunity to develop a speech recognition system for Indian languages. To achieve such aspiring motivation, the research is to develop Marathi speech interface system for talking calculator application.

## 3 Speech Recognition and Feature Extraction Techniques

For the development of Marathi speech activated calculator (MSAC) system, the recognition of speech is necessary. The speech is recognized and then the action is performed. For the speech recognition system speech recording (database creation),



**Fig. 1** Working of speech recognition system

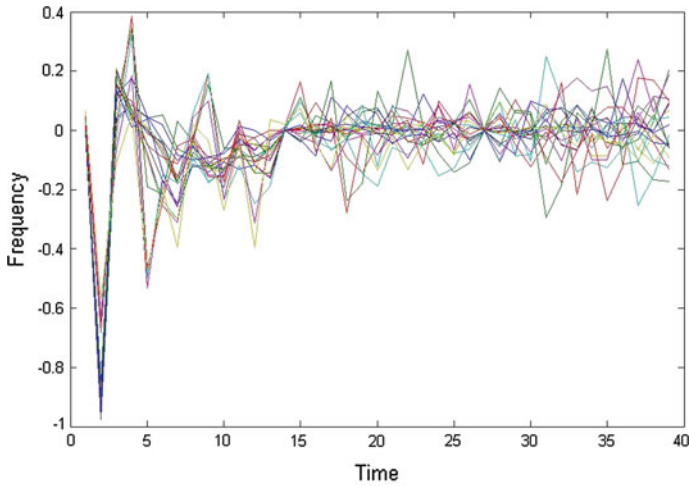
feature extraction, training, classification, and testing are fundamental steps. The step by step flow diagram of speech recognition system is described in Fig. 1.

### 3.1 Feature Extraction Techniques

The MSAC system is trained and tested using three approaches of feature extraction, such as basic feature extraction techniques, fusion approach of feature extraction techniques and Wavelet Decomposed Cepstral Coefficients (WDCC): Proposed Approach. In the basic feature extraction techniques, the Mel Frequency Cepstral Coefficient (MFCC), Linear Discriminative Analysis (LDA), Principal Component Analysis (PCA), LPC, Rasta-PLP Analysis, and Discrete Wavelet Transformation were implemented.

#### (a) *Mel Frequency Cepstral Coefficient*

The enriched literature available on speech recognition, hence reported that the MFCC is most popular and robust technique for feature extraction [17, 18]. The MFCC is based on the known variation of the human ear's critical bandwidth frequencies with filters spaced linearly at low frequencies [19]. In this experiment, we extracted the 13 and 39 features of MFCC. Figure 2 shows the graphical representation of shows 39 Mel Frequency Cepstral coefficients (MFCC) for गणकथंत्त् of first seven frames. The MFCC extracted the basic feature, variation of energy feature.



**Fig. 2** Graphical representation of MFCC 39 coefficients of word गणकयंत्र

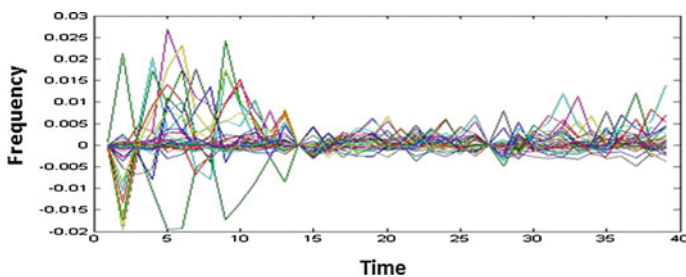
(b) *Linear Discriminative Analysis (LDA)*

LDA algorithm provides better classification compared to principal components analysis [20]. From the literature the Linear Discriminant Analysis (LDA) is commonly used technique for data classification and dimensionality reduction, but in this research we used for feature extraction. Figure 3 shows the graphical representation of extracted LDA feature of speech signal गणकयंत्र for first 10 frames.

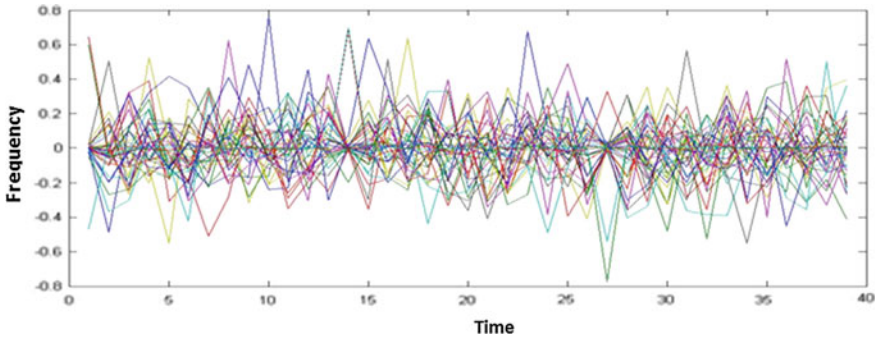
The LDA feature is the combination of the Projection matrix feature, eigenvalues, Mean square representation error, Bias feature, and mean of training data.

(c) *Principal Component Analysis (PCA)*

The principal component analysis is the techniques for classification, but here we are used as feature extraction and dimension reduction. Figure 4 represents the graphical representation of interclass identification of PCA feature.



**Fig. 3** The LDA feature set for the speech signal



**Fig. 4** PCA with class classification

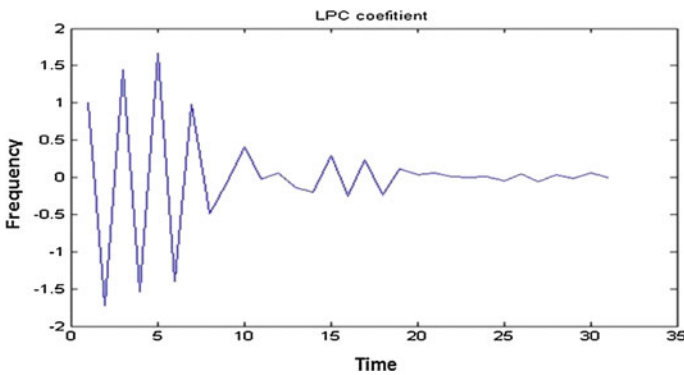
The principal component analysis extracts the 30 feature which is the combinations of projection feature, in class variation, the mean and eigenvalues of the speech signal.

(d) *LPC*

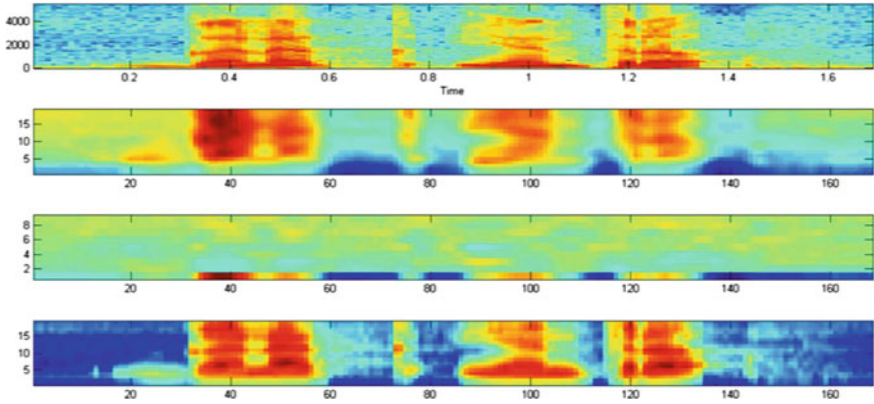
The LPC is one of the robust and dynamic speech analysis techniques. In this research, we used the LPC for feature extraction. The graphical representation of the mean extracted LPC coefficient of speech signal गणकयंत्र is described in Fig. 5. The LPC 7 coefficients contain the Pitch, gain, and duration coefficient parameters of energy.

(e) *Rasta—PLP*

Rasta—PLP techniques is used for feature extraction. This extracted feature is the combination of the graphical band of voice signal. Figure 6 represents the graphical representation of the extracted Rasta PLP coefficients.



**Fig. 5** Mean of extracted LPC coefficient of the speech signal



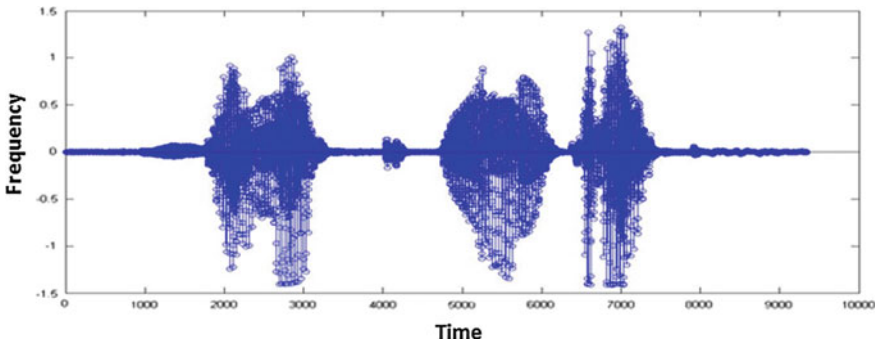
**Fig. 6** RASTA-PLP coefficient of speech signal गणकयंत्र

The Rasta feature contains the pitch, gain and duration of energy and short-term noise coefficients.

(f) *DWT*

The wavelet series is just a sampled version of continuous wavelet transformation and its computation may consume a significant amount of time and resources, depending on the resolution required [21]. The DWT is also used for feature extraction as well as dimension reduction approach. The graphical representation of DWT approximation coefficient is described in Fig. 7.

DWT coefficient is calculated at approximation and detail level. Extracted DWT coefficients include the frequency variation of each frequency band in approximation level and energy variation with time duration in detail coefficients.



**Fig. 7** The extracted approximation coefficient of the speech signal गणकयंत्र

### 3.1.1 Fusion-Based Feature Extraction Techniques

The fusion approach means combination of different techniques. Total 13 MFCC features were extracted and feature vector was formed. The formed feature vector was passed to fusion technique as an input. The detail fusion approach of different techniques with MFCC and their properties is explained in Table 1.

For the fusion-based approach, this research implemented the Mel Frequency Linear Discriminative Analysis (MFLDA), Mel Frequency Principal Component Analysis (MFPCA), Mel Frequency Discrete Wavelet Transformation (MFDWT), Mel Frequency Principal Discrete Wavelet Transformation (MFPDWT), and Mel Frequency Linear Discrete Wavelet Transformation (MFLDWT) fusion approach as feature extraction techniques. Figure 8 represents the graphical representation of extracted MFLDA feature for word गणकयंत्र

### 3.1.2 WDCC: Proposed Approach

In proposed WDCC, the original speech signal is decomposed to second level. The packet coefficient offers different time, frequency representation qualities and consequently potential, for adaptation of the time series phenomenon. This strategy of decomposition offers richest analysis of signal [22–26]. In the WDCC techniques, the original speech signal is decomposed second level. The approximation and detail coefficient is a distinguished output from decomposition step. The DCT operation is performed on horizontal coefficient, which is fused with basic acoustic

**Table 1** Fusion approach with MFCC and their properties

Sr. no	Name of fusion technique	Combination of techniques	Input feature vector	Output feature vector	Properties
1	MFLDA	Fusion of MFCC and LDA	13	02	It is used for dimension reduction
2	MFPCA	Fusion of MFCC and PCA	13	02	It is used for dimension reduction
3	MFDWT	Fusion of MFCC and DWT	13	01	The speed is fast as compare to other techniques
4	MFPDWT	Fusion of MFCC, PCA and DWT	13	01	It is used to reduce time complexity
5	MFLDWT	Fusion of MFCC, LDA and DWT	13	01	It is also used to reduce time complexity



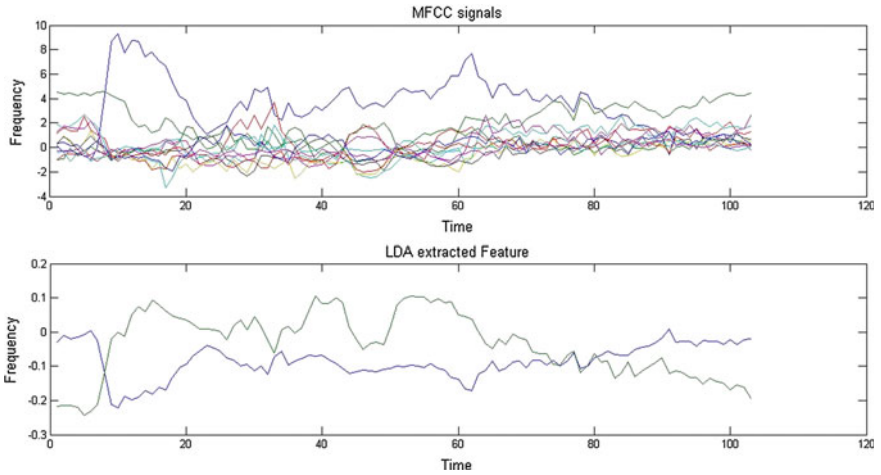


Fig. 8 The fusion approach of MFLDA (MFCC and LDA) of speech signal “गणकयंत्र”

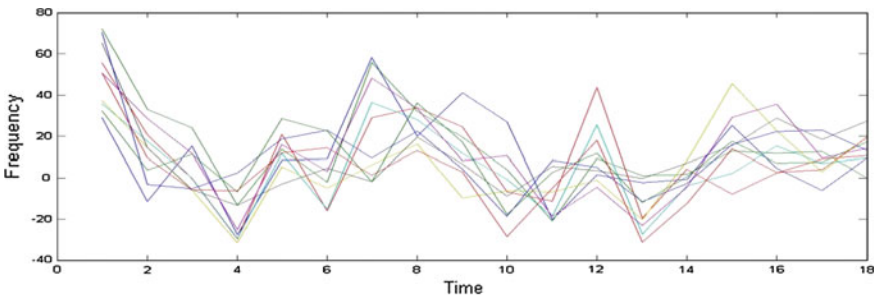


Fig. 9 The WDC extracted 18 features of word “गणकयंत्र”

coefficient are derived to first and second derivation where we got 18, 36, and 54 WDC coefficients. The graphical representation of the extracted WDC 18, 36, and 54 coefficient is shown in Figs. 9, 10 and 11 respectively.

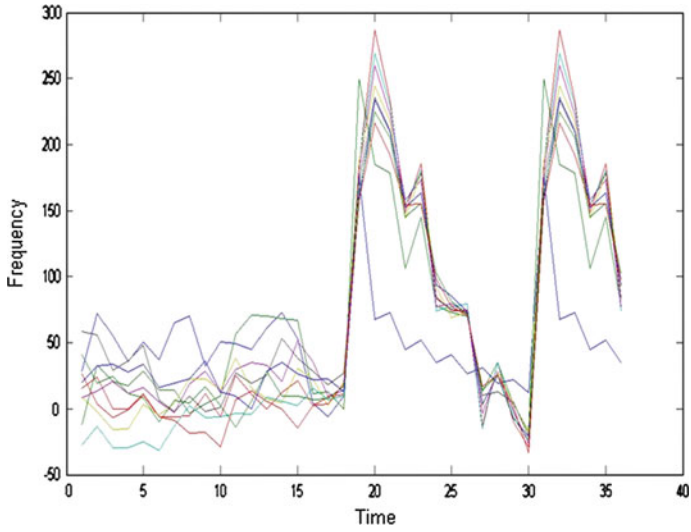


Fig. 10 The WDCC extracted 36 features of word “गणकयंत्र”

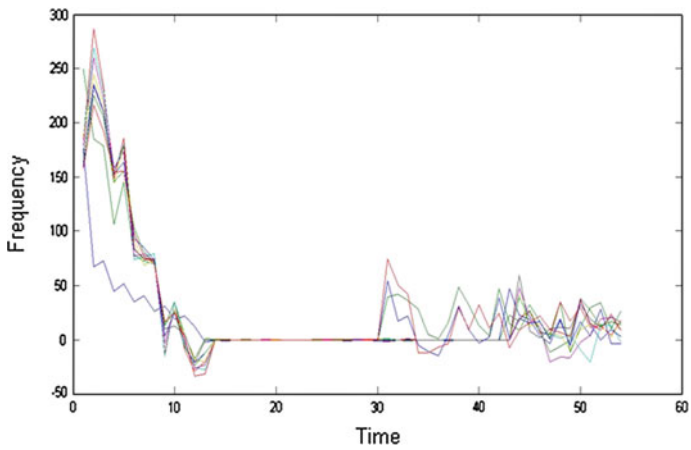


Fig. 11 The WDCC extracted 54 features of word “गणकयंत्र”

## 4 Design of Marathi Speech Interface System

For the speech recognition system speech recording (database creation), feature extraction, training, classification, and testing are fundamental steps.

### 4.1 Database Design

The collection of utterances in the proper manner is called the database. We implemented this prototyping application as a speaker dependent. The total number of words with probability 372 utterance is 20, and the data were collected in 03 session so the overall 22,320/- vocabulary size are collected in the database. The sampling frequency for all recordings was 16,000 Hz at the room temperature and normal humidity. The speech data are collected with the help of microphone realtech and matlab software using the single channel. The preprocessing is done with the help of computerized Speech Laboratory (CSL).

### 4.2 Marathi Speech Activated Calculator (MSAC)

In this research, our objective is to develop MSAC application. Figure 12 describes the basic structural diagram for talking calculator. The voice is recognized and

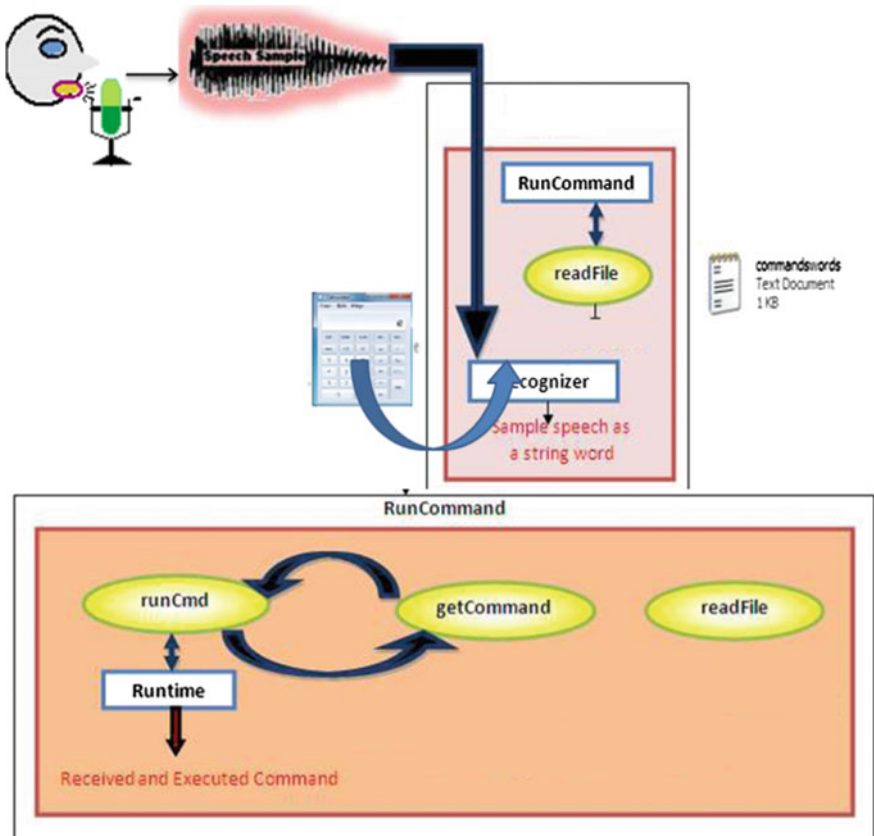


Fig. 12 The basic structural diagram for talking calculator

specific action taken toward voice commands. The MSAC is the speaker-dependent interface system.

## 5 Experimental Analysis

This application deals with defined set of experiments related to calculator applied on the database designed for this research work.

### 5.1 Performance of the Marathi Speech Activated Calculator (MSAC) System

The performance of the system is calculated on the basis of accuracy as well as a real time factor. The real-time factor (RTF) is the time required for recognition in response to the operation. The accuracy is calculated on the basis of confusion matrix in which number of token was passed randomly.

$$\text{Accuracy} = \frac{N - C}{N} * 100$$

where N is a number of token passed and C is a number of token confuse. The RTF is a common metric for computing the speed of an ASR system. If it takes time **P** to process an input of duration **I**, the RTF is defined as

$$\text{RTF} = \frac{P}{I}$$

### 5.2 Training of MSAC System

- MSAC system was trained using individually for the MFCC (13 feature), MFCC (39 feature), LDA, PCA, LPC, Rasta-PLP, DWT techniques, and tested the performance on the basis of the Euclidian distance approach.
- MSAC system was also trained for MFPCA, MFLDA, MFDWT, MFPDWT, and MFLDWT techniques. The performance of MSAC was tested on the basis of the Euclidian distance approach.
- MSAC system was trained using WDCC with 18, 36, and 54 coefficients separately and evaluated for the performance.

### 5.3 Testing of MSAC System

#### (a) Basic Feature Extraction

In this approach, MSAC system was tested using MFCC (13 feature) and MFCC (39 feature), LDA, PCA, LPC, Rasta-PLP, and DWT techniques. The 13 isolated words are used for testing. The performance of these techniques is calculated on the basis of average accuracy and RTF.

- *MFCC based MSAC performance*

The performance of MSAC using MFCC is considered on the basis of 18 and 39 coefficients. Total 13 words were tested for 32 trials. The average performance of MFCC for 18 and 39 coefficients are calculated as 75.78 and 78.03, respectively. The RTF for MFCC 18 and 39 coefficients is the 26 and 38 s, respectively.

- *LDA-based MSAC performance*

Total 13 words were tested for 32 trials. The average performance of LDA is calculated as **67.17 %**. The responding time (RTF) for the action taken is **46 s**.

- *PCA-based MSAC performance*

Total 13 words were tested for 32 trials. The average performance of PCA is calculated as **62.19 %**. The RTF for MSAC using PCA techniques is **38 s**.

- *LPC-based MSAC performance*

Total 13 words were tested for 32 trials. The average performance of LPC is calculated as **61.23 %**. The RTF for recognition and action taken in calculator is **51 s**.

- *Rasta-PLP-based MSAC performance*

Total 13 words were tested for 32 trials. The average performance of Rasta-PLP is calculated as **68.27 %**. The RTF for recognition and action taken in calculator is **48 s**.

- *DWT based MSAC performance*

Total 13 words were tested for 32 trials. The average performance of Rasta-PLP is calculated as **71.02 %**. The responding time for the calculator for performing the action is **32 s**.

The comparative performance of the MSAC with different feature extraction techniques is described in Table 2.

From Table 2, it is observed efficient accuracy is achieved with MFCC for 39 coefficients but RTF is bit increased than MFCC with 13 coefficients. MFCC 13 coefficient proved to be effective in term of accuracy and RTF.

#### (b) Fusion-Based Feature Extraction Techniques

In the fusion feature extraction techniques base MSAC testing, we have tested the system using MFLDA, MFPCA, MFDWT, MFLPDWT, and MFLDWT techniques. We explored fusion approach for system implementation. If the dimension

**Table 2** The performance of MSAC of available feature extraction technique in literature

Sr. no	Technique	Accuracy (%)	RTF (%)
1	PCA	62.19	38
2	LDA	67.17	46
3	LPC	61.23	51
4	Rasta-PLP	68.27	48
5	DWT	71.02	32
6	MFCC (13 feature)	75.78	26
7	MFCC (39 feature)	78.03	38

of the features is reduced without loss of information, this will reduce the RTF and the MFCC provides higher accuracy so we fuse these techniques with MFCC.

- *MFLDA-based MSAC performance*

Total 13 words were tested for 35 trials. The average performance of MFLDA is calculated as **87.90 %**. The responding time for the calculator for performing the action is **22 s**.

- *MFPCA-based MSAC performance*

Total 13 words were tested for 35 trials. The average performance of MFPCA is calculated as **87.12 %**. The responding time for the calculator for performing the action is **20 s**.

- *MFDWT-based MSAC performance*

. Total 13 words were tested for 35 trials. The average performance of MFPCA is calculated as **86.17 %**. The responding time for the calculator for performing the action is **6 s**.

- *MFPDWT-based MSAC performance*

Total 13 words were tested for 35 trials. The average performance of MFPCA is calculated as **89.09 %**. The responding time for the calculator for performing the action is **12 s**.

- *MFLDWT-based MSAC performance*

Total 13 words were tested for 35 trials. The average performance of MFPCA is calculated as **90.12 %**. The responding time for the calculator for performing the action is **10 s**.

The comparative performance of MFLDA, MFPCA, MFDWT, MFPDWT, and MFLDWT are described in Table 3.

The MSAC Application is the real interface application which requires lowest real time factor. From the above results, it is observed that the MFLDWT is proven a higher accuracy than other techniques with acceptable RTF.

**Table 3** Performance of the MSAC system using fusion approach

Sr. no	Fusion approach	Accuracy (%)	RTF
1	MFLDA	87.9	22
2	MFPCA	87.12	20
3	MFDWT	86.17	6
4	MFPDWT	89.09	12
5	MFLDWT	90.12	10

In the above fusion approach, we observed that the wavelet play an important role for the reducing the RTF so we proposed our own techniques on the basis of wavelet basic properties known as WDCC.

(c) *WDCC: A proposed approach*

This wavelet decomposed Cepstral Coefficient is used for basic feature extraction. This technique extracted basic feature, first derivative as well as second derivative. From this technique extracted the 18 basic features, 36 features (18 + First Derivative) and 54 features (18 + First derivative + Second derivatives).

The performance of MSAC is tested on the basis of the proposed WDCC approach. The MSAC system is tested using WDCC 18, 36, and 54 coefficients approach. Total 13 words were tested for 35 trials. The average performance of WDCC with 18, 36, and 54 coefficients is calculated as **94.03, 95.01, and 98.04**, respectively. The responding time for the calculator for performing the action using WDCC 18, 36, and 54 coefficients are 9, 20, and 27 s. The comparative performance of the MSAC system using WDCC technique of different features is described in Table 4.

From the above result, the WDCC with 54 coefficients gives the high accuracy with the RTF is slightly greater than other 18 and 36 coefficient. We observe that WDCC with 18 coefficients provided lowest RTF so we tested the MSAC system using WDCC using 18 coefficients.

For this MSAC interface system, we did 6 h training of the dataset. The testing environment varied the system performance. We test the MSAC interface system real time in the air conditioners in office environment, system processor sound, and group talking environment.

Error Rate in Noisy Environment for MSAC System

This MSAC system is the speaker-dependent system. The experiment is tested in noisy environment. The testing environment varied the system performance. We test the MSAC interface system real time in the following noisy environment. The error rate according to the acquisition environment is presented in Table 5.

**Table 4** Performance of MSAC system using WDCC approach

Sr. no	Technique	Accuracy (%)	RTF
1	WDCC (18 coefficient)	94.03	9
2	WDCC (36 coefficient)	95.01	20
3	WDCC (54 coefficient)	98.04	27

**Table 5** Error rate of MSAC in the different testing environment

Sr. no	Environment	Error rate (%)
1	Air conditioners	13
2	System processor sound	11
3	Group talking environment	10

- **Air conditioners in an office environment** (when the air conditions are started in the office and the sound of the air conditions is mixed with test recorded samples).
- **System Processor sound** (The system processor sound is become maximum and it mixed with the test sample)
- **Group talking environment** (This system tested in real laboratory, number of people talking of each other naturally, they not knowing the system testing background).

#### 5.4 Significance of MSAC

The salient feature of this research as below...

- In the current era of technology, the evolution of speech recognition is done day by day so it is very necessary to bring this technology to the societies in regional language. It is observed that work done in Marathi has not received much more attention. Thus, we have attempted to design and developed Marathi Speech Activated Calculator (MSAC).
- The MSAC system is directly beneficial for society people, where no need of computer literacy and knowledge of English.
- The clustering approach such as PCA, LDA, and LPC was considered for feature extraction, and this gives a new path for speech researcher towards an implementation of real-time application.
- From 1939 to till date, MFCC is one of the robust and dynamic feature extraction techniques; in this experiment, we proposed WDCC feature extraction technique, which gives better performance than MFCC.
- It is very important to adapt research in real-time application; MSAC is the real-time application in Marathi regional language. This will be a chance for Marathwada people to adapt new technology in speech recognition through which they become the part of today's modern technology.

## 6 Conclusion

From the enrich literature, it is observed that MFCC, PCA, LDA, LPC Rasta-PLP, and much more techniques available for feature extraction. Individual techniques have their own limitation. We tried to come up with these limitations using fusion approach. The database for the said application is recorded by standard protocol.



The MSAC system is tested on the basis of individual feature extraction techniques, fusion approach of MFCC and proposed WDCC approach. From the analysis, we observed the following:

- Efficient accuracy is achieved with MFCC for 39 coefficients but RTF is bit increased than MFCC with 13 coefficients. MFCC 13 coefficient proved to be effective in term of accuracy and RTF.
- MFLDWT is proven a higher accuracy than other fused techniques with acceptable RTF.
- WDCC with 54 coefficients gives the high accuracy but the real time factor is slightly greater than other 18 and 36 coefficients.
- WDCC with 18 coefficients provided acceptable accuracy with lowest RTF.
- The testing environment such as air conditioners in office, system processor sound and group talking environment varies the performance of MSAC.

From the above study the author recommended the WDCC feature extraction techniques are robust and dynamic as compare to MFCC, LPC, Rasta, LDA, and PCA.

## References

1. A review on speech recognition technique. *Int. J. Comput. Appl.* **10**(3), 0975–8887 (2010)
2. Picheny, M.: Large vocabulary speech recognition **35**(4):42–50 (2002)
3. Arokia Raj, A., Susmitha, R.C.: A voice interface for the visually impaired. In: 3rd International Conference: Sciences of Electronic, Technologies of Information and Telecommunications March 27–31, Tunisia (2005)
4. Roux, J.C., Botha, E.C., Du Preez, J.A.: Developing a multilingual telephone based information system in African languages. In: Proceedings of the Second International Language Resources and Evaluation Conference, no. 2, pp. 975–980. ELRA, Athens (2000)
5. Robertson, J., Wong, Y.T., Chung, C., Kim, D.K.: Automatic speech recognition for generalized time based media retrieval and indexing. In: Proceedings of the Sixth ACM International Conference on Multimedia, pp. 241–246. Bristol (1998)
6. Scan soft: Embedded speech solutions. <http://www.speechworks.com/> (2004). Accessed 25 Jan 2013
7. Kandasamy, S.: Speech recognition systems. *SURPRISE J.* **1**(1) (1995)
8. Dusan, S., Rabiner, L.R.: On integrating insights from human speech perception into automatic speech recognition. In: Proceedings of INTERSPEECH 2005. Lisbon (2005)
9. Shrawankar, U., Thakare, V.: Speech user interface for computer based education system. In: International Conference on Signal and Image Processing (ICSIP), pp. 148–152 (2010) (15–17 Dec)
10. Alt, F.L., Rubino, M., Yovits, M.C.: *Advances in Computers*, pp. 165–230. Academic Press, New York
11. Rebman Jr., C.M., Aiken, M.W., Cegielski, C.G.: *Speech Recognition in the Human-Computer Interface*, vol. 40, Issue 6, pp. 509–519, Information & Management. Elsevier (2003)
12. Furui, S.: 50 Years of progress in speech and speaker recognition research. *ECTI Trans. Comput. Inf. Technol.* **1**(2) (2005)

13. Nehe, N.S., Holambe, R.S.: New feature extraction techniques for Marathi digit recognition. *Int. J. Recent Trends Eng.* **2**(2) (2009)
14. Bhosale, R.S.: Enhanced speech recognition using ADAG SVM approach. *Int. J. Emerg. Trends Technol. Comput. Sci. (IJETTCS)* **1**(4) (2012)
15. Anumanchipalli, G., Chitturi, R., Joshi, S., Kumar, R., Singh, S.P., Sitaram, R.N.V., Kishore, S.P.: Development of indian language speech databases for large vocabulary speech recognition systems. In: *Proceedings of International Conference on Speech and Computer (SPECOM)*. Patras (2005)
16. Neti, C., Rajput, N., Verma, A.: A large vocabulary continuous speech recognition system for Hindi. In: *Proceedings of the National conference on Communications*, pp. 366–370. Mumbai (2002)
17. Gawali, B.W., Gaikwad, S., Yannawar, P., Mehrotra, S.C.: Marathi Isolated Word Recognition System using MFCC and DTW Features. *ACEEE* (2010)
18. Chakraborty, K., Talele, A., Upadhy, S.: Voice recognition using MFCC algorithm. *Int. J. Innovative Res. Adv. Eng. (IJIRAE)* **1**(10) (2014). ISSN: 2349-2163
19. Patel, K., Prasad, R.K.: Speech recognition and verification using MFCC & VQ. *Int. J. Emerg. Sci. Eng. (IJESE)* **1**(7) (2013). ISSN: 2319–6378
20. Oh-Wook Kwon, Chan, K., Lee, T.-W.: Speech feature analysis using variational bayesian PCA. *IEEE Signal Process. Lett.* **10**, 137–140 (2003)
21. Gaikwad, S., Gawali, B., Mehrotra, S.C.: Novel Approach Based Feature Extraction For Marathi Continuous Speech Recognition, pp. 795–804. *ACM Digital Library*, New York (2012). ISBN: 978-1-4503-1196-0/2012
22. Hermansky, H., Morgan, N.: RASTA processing of speech. *IEEE Trans. Speech Audio Process.* **2**, 578–589 (1994). doi:[10.1109/89.326616](https://doi.org/10.1109/89.326616)
23. Ali, H., Ahmad, N., Zhou, X., Iqbal, K., Muhammad Ali, S.: DWT features performance analysis for automatic speech recognition of Urdu. *SpringerPlus* **3**:204 (2014) doi:[10.1186/2193-1801-3-204](https://doi.org/10.1186/2193-1801-3-204)
24. Tiwari, A., Zadgaonkar, A.S.: Debauchee’s wavelet analysis of speech signal of different speakers for similar speech set. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* **4**(8) (2014)
25. Mallat, S.: *A Wavelet Tour of Signal Processing*. Academic Press (1998)
26. Chan, Y.T.: *Wavelet Basics*. Kulwer Academic Publications (1995)