

DESIGN AND DEVELOPMENT OF SPEECH DATABASE FOR DIGIT AND VOWELS IN SANSKRIT LANGUAGE AND RECOGNITION SYSTEM USING MACHINE LEARNING

Abstract

The study of machine voice processing is called speech recognition. The synthesis component may be referred to as speech production. Speech recognition is a prominent and active field of study that translates human speech into computer-recognizable words. It often entails pattern extraction from digitized audio samples and representation of those patterns using the proper data model. The content of these patterns is then determined by comparing them to one another using mathematical processes. The focus of this research is speech recognition of isolated words. The purpose of this study is to identify Sanskrit isolated words uttered by various male and female non-native Sanskrit speakers from Maharashtra's Aurangabad District. The MFCC, or Mel Frequency Cepstral Coefficients In order to test the samples, features from and Support Vector Machine (SVM) of the audio signal have been used. Confusion Matrix is used for classification

Author

Ashwini Shivaji Ganakwar
M. Phil Student
Department of Computer Science and
Information Technology
Dr. Babasaheb Ambedkar
Marathwada University
Aurangabad, Maharashtra, India.
ashwiniganakwar22@gmail.com

I. INTRODUCTION

Speech stands out as the foremost and instinctive mode of human interaction and communication [1]. It holds a central role in human engagement, enabling natural and effective communication among individuals. Globally, speech represents a predominant means of interaction [1]. The significance of speech lies in its facilitation of seamless interpersonal communication, rendering it a pivotal element of human activity. Within the realm of technology, speech recognition pertains to the process of transcribing spoken language into a machine-readable format. Essentially, speech recognition involves conversing with a computer and having it accurately interpret spoken words [2]. This capacity for speech interaction has the potential to revolutionize human-computer communication.

The aspiration to create machines capable of comprehending and communicating akin to humans has been a longstanding ambition. In this pursuit, researchers have strived to construct systems that analyze and categorize speech signals. The applications of such advancements span across diverse sectors, including agriculture, healthcare, and government services, all of which could substantially benefit from computers proficient in understanding spoken language. Presently, only a fraction of those who comprehend a given language can access the vast expanse of information within the digital sphere. Language technology serves as a solution by offering user-friendly interfaces that enable the widespread dissemination of digital content, bridging linguistic barriers [3]. Automatic Speech Recognition (ASR) systems have achieved commendable accuracy levels, endowing them with potential utility in various sectors where they can streamline daily operations. One of the domains benefiting from these advances is speech-based Human-Computer Interaction (HCI), especially in scenarios where conventional keyboards prove inadequate. This avenue of exploration holds promise, particularly when seeking a natural language communication interface [3]. Researchers are actively innovating new interfaces for interaction with computer systems, harnessing the potential of speech as a means of engagement [3]. This research delves into the design and creation of databases containing isolated digits and vowels for the Sanskrit language. The primary requisite is a corpus of textual data captured from diverse Sanskrit speakers. While substantial work has been undertaken for other languages, the domain of Sanskrit language has received comparatively limited attention. This study endeavors to fill this gap by amassing a comprehensive database of isolated speech samples and investigating the nuances of spoken Sanskrit.

- 1. About Sanskrit Language:** Sanskrit is an ancient Indo-Aryan language that holds a significant place in the linguistic and cultural history of India. It is considered one of the oldest languages in the world and is often referred to as the "language of the gods" due to its association with religious and philosophical texts in Hinduism, Buddhism, and Jainism. Here is some detailed information about the Sanskrit language:
- 2. Historical Significance:** Sanskrit has a rich historical and cultural legacy. It is the liturgical language of many religious texts, including the Vedas, Upanishads, Puranas, and various other sacred scriptures. These texts are foundational to the religious and philosophical traditions of India and have played a vital role in shaping the spiritual and intellectual landscape of the subcontinent.
- 3. Linguistic Features:** Sanskrit is known for its complex grammatical structure and precision. It has a highly evolved system of declensions, conjugations, and verb forms. Its grammar is meticulously described in ancient texts like Panini's "Ashtadhyayi," which is one of the earliest known grammatical treatises in the world.

The language is categorized into different periods based on its development and usage:

- 4. Vedic Sanskrit:** The language of the ancient Hindu scriptures, known as the Vedas.
- 5. Classical Sanskrit:** The language of classical literature, including epic poems like the Ramayana and Mahabharata, as well as philosophical and dramatic works.
- 6. Medieval Sanskrit:** The language evolved during the medieval period and witnessed the composition of various literary, scientific, and philosophical texts.
- 7. Modern Sanskrit:** Though not widely spoken as a native language, Sanskrit continues to be used for academic and scholarly purposes. Some contemporary writers also compose poetry and literature in Sanskrit.

Contribution to Other Languages: Sanskrit has had a profound influence on many languages in the Indian subcontinent. Many modern Indo-Aryan

languages, such as Hindi, Bengali, and Marathi, have evolved from Sanskrit and share vocabulary, grammatical features, and linguistic patterns with it.

8. Revival and Preservation: Efforts have been made to revive and promote the study of Sanskrit in modern times. Some universities and institutions offer courses in Sanskrit, and there are ongoing discussions about integrating Sanskrit into technological domains, including natural language processing and machine learning, due to its structured grammar and precision.

9. Challenges and Future: While Sanskrit remains a cherished part of India's cultural heritage, it is not spoken as a primary language by a large population. Its complex grammar and limited contemporary usage pose challenges to its widespread adoption. However, its influence on various aspects of Indian culture, religion, philosophy, and literature ensures that it continues to hold a place of reverence and importance.

In conclusion, Sanskrit is a language that carries a profound historical, cultural, and linguistic legacy. Its impact on literature, philosophy, and religion is immeasurable, making it a subject of fascination and study for scholars, linguists, and enthusiasts around the world.

10. Objective of research:

- To design and develop text corpora for Sanskrit Language.
- To collect speech samples from digit and vowels.
- To Recognition of prosody features of database.
- To design and develop the recognition system for Sanskrit Language.

II. LITERATURE SURVEY

Sr. No.	Author	Techniques	Conclusion/Observation
1	Pukhraj P. Shrishrimal Ratnadeep R. Deshmukh Vishal B. Waghmare	LDC-IL	This essay examines the Hindi, Tamil, Telugu, and Bengali varieties of the Indian language. Little has been done, and little is being done, to support the Marathi language.
2	Potale shubham , Kharpude Pratik, Patil rahul Ajay	MFCC, LPCC	After reading this paper, I came to the conclusion that by using the MFCC approach, I had finished the

	kumar gupta		work of the isolated word.
3	Yogesh K. Gedam, Sujata S. Magare, Amrapali C. Dabhade, Ratnadeep R. Deshmukh	DTW,MFC C	The primary goal of their effort is to identify solitary speech using the approaches listed below. DTW and MFCC
4	Devyani S. Kulkarni Ratnadeep R. Deshmukh Pukhraj P. Shrishrimal	DFT, LPC, DFT, Single Chanel Enhanceme nt Techniques, NSS, ANC	This research examines several noise sources and methods for removing them with the goal of improving voice signal quality. and numerous strategies are employed in the study.
5	V. K. Kale, 2R.R. Deshmukh, G. B. Janvale	MFCC, LPC, DTW, HMM	Work has been done to identify English isolated words and determine how native and non-native speakers differ.
6	Ashok Kumar, Vikas Mittal	MFCC, LPC, LPL, DTW, HMM	Different techniques are used in this paper, and speech recognition is used in this paper.
7	Yogesh K. Gedam, Paras V. Kothare, Ratnadeep R. Deshmukh	ANN, PRAAT	This document includes the data gathered for creating the ASR for Marathi numerals for agricultural use.
8	Suman K. Saksamudre, P.P. Shrishrimal, R.R. Deshmukh	ANN, HMM	In this research, many strategies were employed, and with their aid, it was possible to match the patterns for improved results.
9	ShaikhNaziya S. , R. R. Deshmukh	LPC, HMM	They achieve a result of 100% on this paper.
10	Pukhraj P. Shrishirmal, Ratnadeep R. Deshmukh, Vishal B. Waghmare, Sushma Borade, Pooja V. Ganesh	PRAAT	The purpose of this work is to construct and develop a Marathi language speech corpus from the Marathwada region. One isolated word speech database and two isolated speech databases for the agriculture domain

	B. Janvale		
11	Mr.V.K.Kale, Dr.R.R.Deshmukh, Dr.G.B.Janvale, Mr.V.B.Waghmare , Mr.P.P.Shrishrimal	LPC, MFCC.	Through confusion matrix, they were able to recognise MFCCs with a 95.75% accuracy and LPCs with a 61.40% accuracy utilising these methodologies.
12	Pukhraj P. Shrishrimal Ratnadeep R. Deshmukh Vishal B. Waghmare	PCA	The creation of isolated words for agricultural purposes from the Aurangabad District of Maharashtra can use the speech database that has been produced.
13	V. B. Waghmare R. R. Deshmukh G. B. Janvale	Spectra l Subtraction Method, PRAAT	This paper makes use of a database of various emotive speech in several languages.
14	Aaron M. Oirere , Ganesh B. Janvale Ratnadeep R. Deshmukh	LPC, MFCC and SVM	In this study, features from three databases—agriculture, isolated words, and words and sentences uttered by native and non-native speakers—were collected. These features were then extracted using MFCC and LPC.
15	Pratik K. Kurzekar ,Ratnadeep R. Deshmukh , Vishal B. Waghmare , Pokhran P. Shrishrimal	MFCC, LPC, DTW	They developed (swahil) using an isolated Swahili language numeric database. Covering the fundamental phonetics of the Swahili language will be helpful for both native and non-native speakers. The work can be expanded to include problematic Swahili speech, as the current effort only covers numbers 0 to 9.
16	Aaron M. Oirere, Ratnadeep R. Deshmukh, Pukhraj P. Shrishrimal	NLP	For models 1 and 2, recognition rates of 72.96% and 76.96% were attained. The simplest approach for creating a model based on decision boundaries is the k-nn algorithm.

- 1. Methodology:** The goal of this end ever is to identify isolated spoken words in the Sanskrit language based on provided input. To achieve this system, we employ the components outlined in the basic block diagram below.

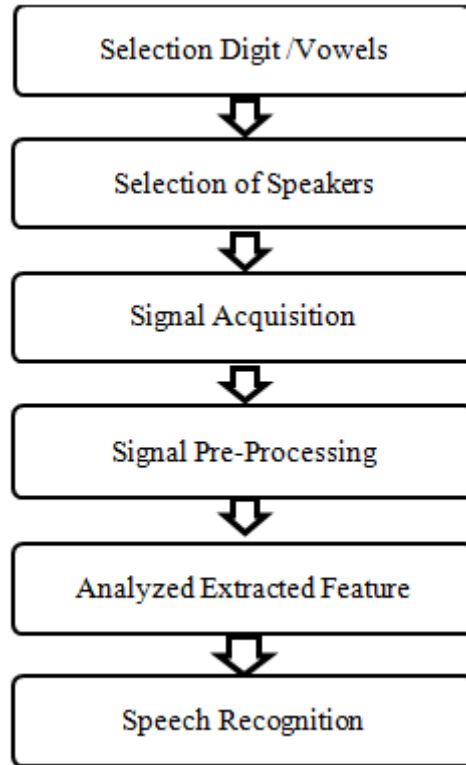


Figure: Work Flow of Proposed System

The above fig.1 the process work is divided into several steps, we begin with the selection of text corpora i.e. digit or vowels sample we have collect from the Sanskrit speakers. Then we select the speak of Sanskrit speakers. Once the speech samples are collected, we will be performing pre-processing. The next step is to analyze the extracted features of Sanskrit speakers. The final step is to recognize the speech of speakers for Sanskrit language.

- 2. Data Collection Process|:** The procedures for developing speech corpora are discussed at this level. Having selected the recording medium, the data is captured using a high-quality microphone and a laptop equipped with PRAAT for recording speech signals.

- **Speaker Choice:** Non-native speakers of Sanskrit will be used to collect the speech data. The selected speakers will be provided by the District of Aurangabad [7,8,9]. They would find it difficult to read or speak Sanskrit.

Table 1: Sanskrit Digits

Number	0	1	2	3	4	5	6	7	8	9
Sanskrit word	शून्य	एकम्	द्वे	त्रीणि	चत्वारि	पञ्च	षट्	सप्त	अष्ट	नव
IPA	Śūnya	Ēkam	Dvī	Trī	chatvari	Pañca	śaṣ	Saptá	aṣṭá	Náva

Table 2: Sanskrit Vowels

Letter	अ	आ	इ	ई	उ	ऊ	ऋ	ॠ	ऌ	ॡ	ए	ऐ	ओ	औ	अं	अँ	अः
Diacritic	◌	◌ा	◌ि	◌ी	◌ु	◌ू	◌ृ	◌ॄ	◌ॢ	◌ॣ	◌े	◌ै	◌ो	◌ौ	◌ं	◌ँ	◌ः
Name	A	Ā	I	Ī	U	Ū	r̥	r̄	l̥	l̄	E	Ai	O	Au	aṅ	aṁ	aḥ
IPA	/ʌ/	/a:/	/i/	/i:/	/u/	/u:/	/r/	/r:/	/l/	/l:/	/e:/	/a:i/	/o/	/a:u/	/aŋ/	/aṁ/	/əh/

III. SPEECH DATA GATHERING

To record the speech, we utilised the PRAAT Software. by recording the speech samples using a Headset Quantum QHM862 USB Stereo headset. After choosing the recording medium, the data is a laptop with PRAAT for capturing speech signals and a high-quality microphone were used to record the speech signals [19].

- **Step 1:** Selected speakers were questioned about any issues they experienced reading or reciting the isolated Sanskrit words.
- **Step 2:** Speakers were provided a brief tutorial on how to use the headset and when to talk [20].
- **Step 3:** The sample frequency for a Mono sound type was set to 16000 Hz.
- **Step 4:** After asking the speaker for each syllable, recording the sample and saving As a.wav file, it.

For each of the 1500 utterances that were captured from the speaker, step 4 was repeated. Each step was carried out once for each of the 20 speakers.

- 1. Data Gathering Statistics:** The speech data was contributed by 20 speakers. Each speaker has to say three utterance totalling 25 words. Each speaker has to pronounce seventy-five words. A total of 1500 word utterances have been recorded. 1500 utterances have so far been collected from 20 speakers, comprising 10 men and 10 women [21, 22].

The environment is important. The graphical user interface of PRAAT is its key advantage [24]. Numerous facilities are available, such as those for spectrum analysis, pitch analysis, voice analysis, format analysis, intensity analysis, and more [25]. In addition, PRAAT has the capacity to perform general analysis (waveform, intensity, spectrogram, pitch, duration).

Table 3: Information about Data Collection Procedure

Process	Description
Language	Sanskrit
Tools	Praat, Headset Quantum QHM862 USB Stereo
Total Number of speakers	10 Male and 10 Female
Frequency	16000 Hz
Total Utterances	1500

Table 4: Database of Speech Metadata

Process	Description
Total Number of Chosen Words	25 Utterances Recorded
Utterances Recorded	Three times each word is said.
Utterance total per Speaker	75

Complete Speaker	20
(Male) Speaker	10
(Female) Speaker	10
Number of Male speakers overall	750
Number of Female speakers overall	750
Complete Utterances	1500
Software Used for Recording	PRAAT
Tool	Headset Quantum QHM862 USB Stereo
Recording Frequency	16000Hz

IV. PERFORMANCE ANALYSIS

1. Mel Frequency Cepstral Coefficient (MFCC): Mel Frequency Cepstral Coefficient, often known as MFCC, is the most well-known and frequently used feature extraction technique in the speech domain. The human auditory perception system is the foundation of the MFCC. Extraction of MFCC features from audio signals. These features are defined from the 13 MFCC features vectors, which are extracted from the “.wav” file of the input speech signal. Extraction of MFCC features from audio signals. These works take the peak loudness across all channels and will be used to calculate the MFCC [4]. The outcome can be different from the individual MFCC calculation for each channel.

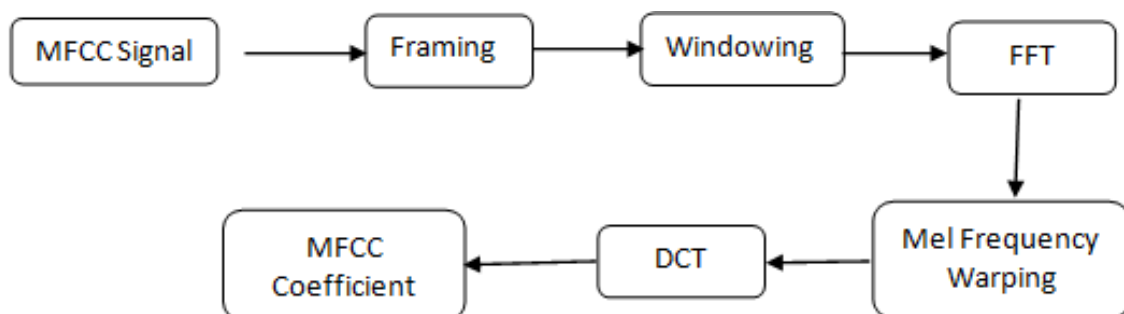


Figure 2: MFCC Feature Extraction

- **Pre-processing:** A pre-emphasis filter is used to pre-process the incoming voice signal, boosting higher frequencies while cutting lower frequencies. As a result, the signal-to-noise ratio is improved.
 - **Frame blocking:** The pre-processed voice stream is split into brief frames that typically last 20 to 30 ms and have a 50% overlap.
 - **Windowing:** To lessen spectral leakage and smooth the frame edges, each frame is multiplied using a window function, such as the Hamming window.
 - **Fast Fourier Transform (FFT):** To obtain the magnitude spectrum, the windowed frames are transformed to the frequency domain using the FFT technique.
 - **Mel Frequency Filter Bank:** In order to establish a filter bank, a number of peaks that are evenly spaced on the Mel-scale are calculated, and once they have been transformed back to normal frequency scale, they are employed as peaks for the filter banks [4,5].
 - **Cepstral analysis:** The Mel-frequency Cepstral Coefficients (MFCCs) are obtained by applying the Discrete Cosine Transform (DCT) to the Mel-scaled spectrum in the Cepstral domain. Typically, the first 12 MFCCs are used, which capture the most important spectral features.
 - **Energy coefficients:** Finally, the log of the energy of each frame is computed, and the delta and double delta coefficients are also computed for the energy.
 - **Discrete Cosine Transform:** To obtain the mel- cepstrum coefficients, (DCT) is being used. There are 24 coefficients in a frame, however only 13 have been chosen for the recognition system. Only 13 of the 24 Mel Cepstral coefficients have been chosen for the recognition system.
- 2. Support Vector Machine (SVM):** A well-known machine learning approach called Support Vector Machine (SVM) is utilised for classification and regression problems. It locates the ideal hyperplane that maximises the margin between several classes of data points and best separates them. SVM is very useful since it can handle complicated data distributions and produce good classification results. In order to find a hyperplane that successfully

separates the classes, the method first transforms the input data into a higher-dimensional space. SVM was essential to the success of our word recognition studies. Word recognition includes speech recognition software's ability to decode spoken words. For this challenge, our strategy made use of the Support Vector Machine algorithm's strength. We meticulously placed each unique learned dataset into the frames, enabling these Features and Test Features to be used as inputs by the SVM Algorithm. SVM techniques fall under a category of supervised learning methods that computer scientists employ to analyze data and discern patterns within it. SVM stands out due to its robust classification performance in comparison to other classifiers. Its capacity to effectively handle complex patterns, even with limited training data, makes it a popular choice for tasks like pattern recognition and classification. Once we extracted features from all files, we harnessed SVM to assess and quantify their similarity. This involved computing the minimum distance between the features of each test file and the corresponding training files. This process enabled us to accurately evaluate how well the test files aligned with the training data, facilitating precise word recognition [12].

V. RESULTS AND CONCLUSION

This study focuses on the comparison of performance between Mel Frequency Cepstral Coefficient (MFCC) with 13 coefficients and Support Vector Machine (SVM). The experimental speech data comprised isolated Sanskrit words spoken by various speakers. For the feature extraction phase, we opted for the utilization of Mel Frequency Cepstral Coefficient (MFCC). Our pursuit of enhancing speaker-level accuracy led us to integrate diverse techniques and additional audio features, which notably contributed to improved recognition and differentiation of speakers.

The investigation centers on assessing the efficacy of SVM and MFCC (with 13 coefficients) by analysing isolated Sanskrit words spoken by female speakers. The chosen feature extraction method, Mel Frequency Cepstral Coefficient (MFCC), was enriched with various strategies and supplementary audio attributes to augment system accuracy at the speaker level. This augmentation facilitated more proficient speaker differentiation and recognition.

Confusion Matrix

		Predicted	
		Negative	Positive
Actual	Negative	a	b
	Positive	c	d

1. The Accuracy (AC) calculates the proportion of correctly predicted events. It is calculated using the formulas.

$$AC = \frac{a+d}{a+b+c+d}$$

2. The Recall or True Positive Rate (TP) is the percentage of positive cases that were correctly identified

$$TP = \frac{d}{c+d}$$

3. Precision: Precision displays the proportion of accurately anticipated cases that had positive result.

$$Precision = \frac{TP}{TP + FP}$$

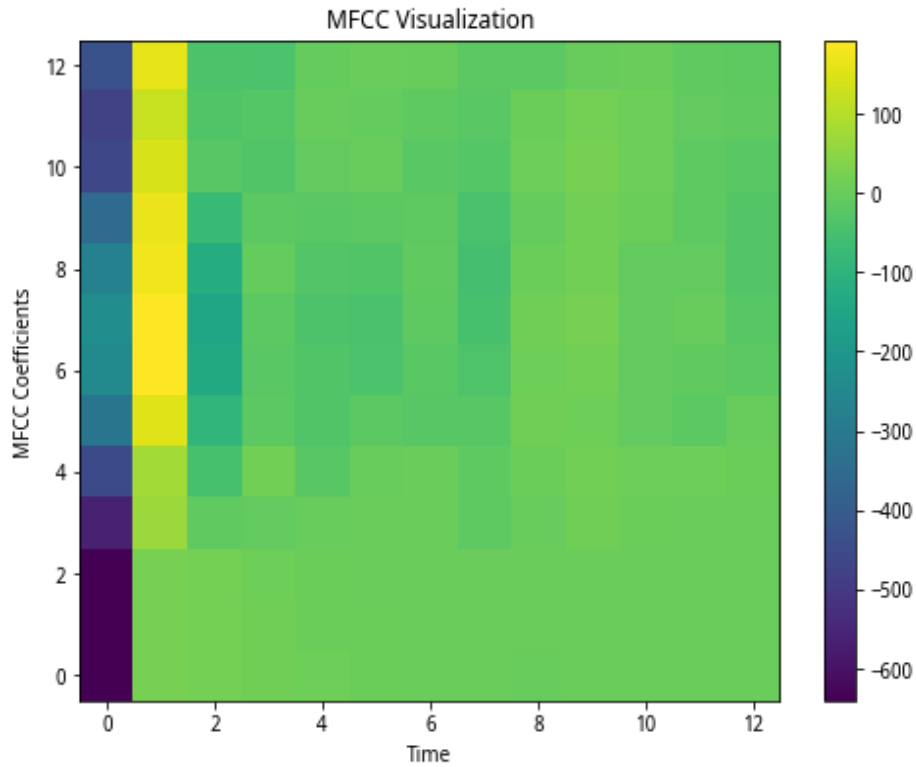
4. Recall: Recall is a significant statistic when False Negative exceeds False Positive.

$$Recall = \frac{TP}{TP + FN}$$

5. F1-Score: The harmonic mean of precision and recall is represented by the F1 score, which ranges from 0 to 1.

$$F1 - score = \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}}$$

MFCC Coefficients



GUI Design & Development

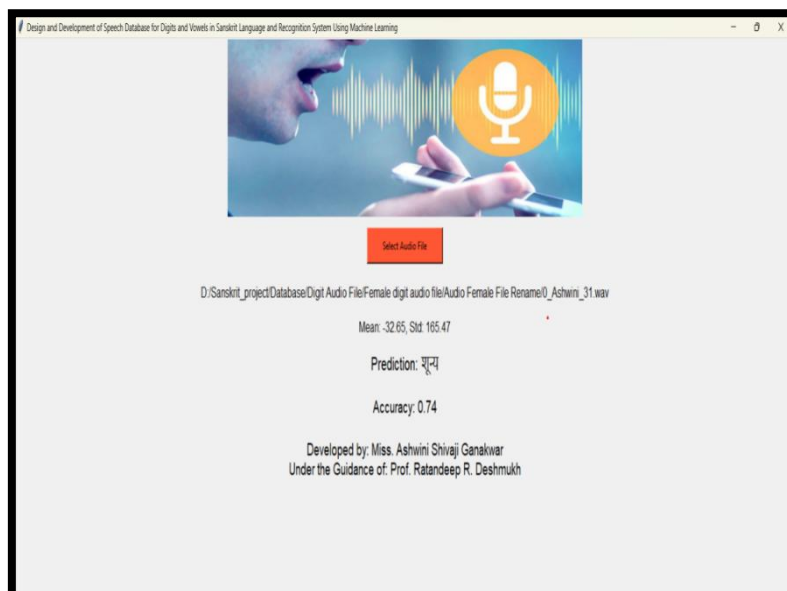


Figure 3: GUI for the Design and Development of speech Database for Sanskrit Language.

VII. MFCC COMPARATIVE RESULT

Table 4: Extracted MFCC Features for Female: Sample- word: "Ekam"(One)

Table Extracted MFCC Features for Female: Sample- word: "Ekam"(One)											
Feature and Coefficient	Frame1	Frame2	Frame3	Frame4	Frame5	Frame6	Frame7	Frame8	Frame9	Frame10	Frame11
C1	-8.37146	-7.55289	-7.57788	-7.59971	-7.56086	-7.86501	-8.36963	-8.44283	-8.46554	-9.66241	-9.1689
C2	-9.70858	-1.17584	-8.71688	-9.21673	-8.49592	-6.26343	-6.10146	-8.98199	-9.10916	-7.29148	-8.04496
C3	2.022595	-2.45494	-1.70734	-1.87292	-8.49592	-2.0769	-1.37353	-1.8339	-1.52864	-3.04711	-3.73716
C4	9.696576	9.395842	1.000102	8.080779	2.762253	-3.09036	4.993647	7.622424	1.730026	1.405737	-6.18979
C5	4.39945	-5.49798	-5.33702	-5.7508	-5.47549	-5.13076	-4.73302	-3.94642	-2.39153	9.977631	6.214581
C6	-6.15067	-5.95755	-5.26554	-4.6221	-4.89887	-5.26355	-4.93902	-5.8345	-5.29597	-1.9182	-2.8958
C7	1.829157	1.886384	1.946486	1.489107	1.247529	1.163114	3.460435	-5.70636	5.765854	-3.56011	-1.49084
C8	-1.17299	-1.35675	3.118666	1.298936	6.557125	-7.65297	-1.85566	-1.94203	-1.99834	-1.06615	-5.58765
C9	-1.60922	-1.71804	-1.72999	-2.13168	-2.40607	-2.27684	-2.5694	-2.92982	-3.96494	-8.25612	3.250899
C10	1.233601	1.155851	1.387194	6.958546	1.307372	1.451853	1.888897	9.42452	3.394384	-8.43262	-1.15005
C11	-5.48966	-5.0665	-8.38863	1.662038	-8.05424	-9.9788	-4.97413	-1.38811	-2.7196	-2.6052	-9.27334
C12	-1.83007	-2.49985	-1.64704	-2.0228	-1.71428	-1.54476	-2.34158	-2.391	-1.67922	-6.66867	1.420725

Table 5: Extracted MFCC Features for Male: Sample- word: " Ekam" (One)

Table Extracted MFCC Features for Male: Sample- word: " Ekam"(One)											
	Frame1	Frame2	Frame3	Frame4	Frame5	Frame6	Frame7	Frame8	Frame9	Frame10	Frame11
C1	-9.32E+00	-9.14E+00	-8.72E+00	-8.63E+00	-8.54E+00	-8.69E+00	-9.61E+00	-1.10E+01	-8.33E+00	-8.49E+00	-8.48E+00
C2	6.96E+00	7.18E+00	8.03E+00	6.69E+00	6.76E+00	4.90E+00	2.41E+00	-2.22E+00	-7.20E+00	-6.88E+00	3.16E+00
C3	4.16E+00	9.55E-01	-1.44E+00	3.10E+00	2.78E+00	1.67E+00	-1.98E+00	-1.87E+01	-4.23E+01	-4.31E+01	3.16E+00
C4	1.95E+01	1.82E+01	1.59E+01	1.52E+01	1.82E+01	2.59E+01	2.35E+01	6.36E+00	8.42E-01	-1.14E+01	-1.59E+01
C5	-7.92E+00	-9.60E+00	-1.03E+01	-9.94E+00	-6.54E+00	-8.59E+00	-4.71E+00	1.55E+01	2.88E+00	-3.16E+00	-1.12E+01
C6	-8.10E+01	-7.10E+01	-7.25E+01	-7.14E+01	-7.21E+01	-6.96E+01	-6.28E+01	-1.12E+01	3.20E+00	3.88E+00	2.19E+00
C7	-1.85E+01	-2.49E+01	-2.97E+01	-3.68E+01	-2.77E+01	-2.39E+01	-1.50E+01	-7.04E-01	-2.95E+00	-1.81E+00	-2.01E+00
C8	-6.30E+00	-1.44E+01	-9.27E+00	-5.64E+00	-5.55E+00	-7.56E+00	-8.83E-01	-1.89E+01	-1.74E+01	-2.06E+01	-2.09E+01
C9	-1.63E+01	-1.02E+01	-1.87E+01	-1.28E+01	-1.54E+01	-1.97E+01	-3.08E-01	7.15E+00	-1.74E+01	1.77E+01	1.36E+01
C10	-1.06E+01	-1.40E+01	-1.41E+01	-6.20E+00	-2.73E+00	-3.62E+00	-7.22E+00	-3.16E+00	4.73E+00	7.97E+00	1.24E+01
C11	7.93E+00	2.61E+00	6.31E+00	-1.01E+00	1.65E+00	1.17E+01	8.59E+00	7.39E+00	1.13E+01	1.40E+01	1.24E+01
C12	-8.33E+00	-1.26E+01	-1.55E+01	-1.82E+01	-7.65E+00	-5.25E+00	-3.82E+00	-1.11E+01	-9.91E+00	-1.22E+01	-1.08E+01

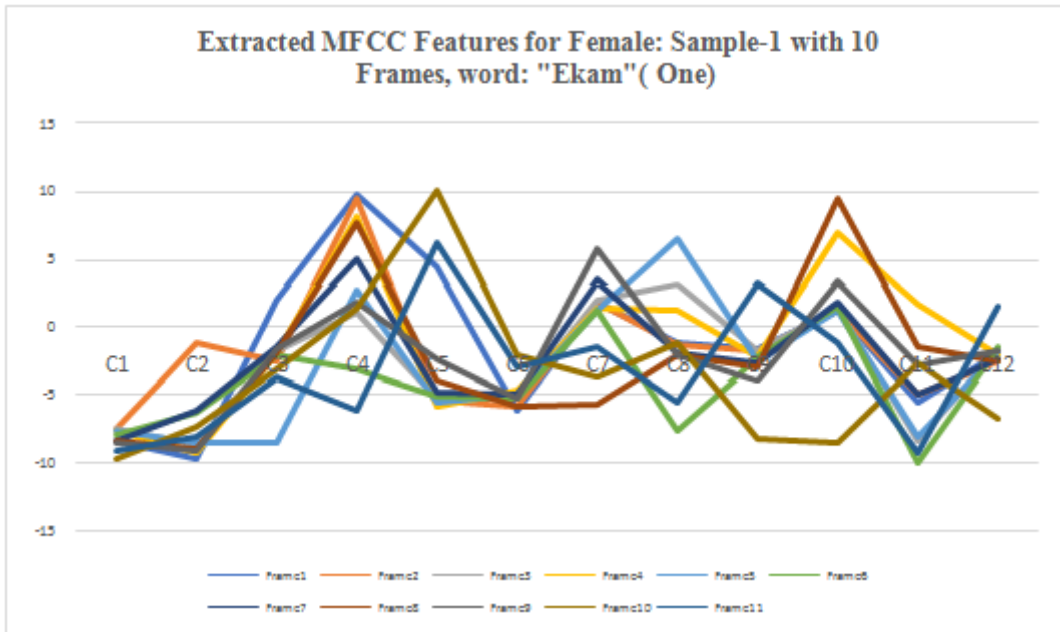


Figure 4: Extracted MFCC Features for Female: Sample word: "Ekam"(One)

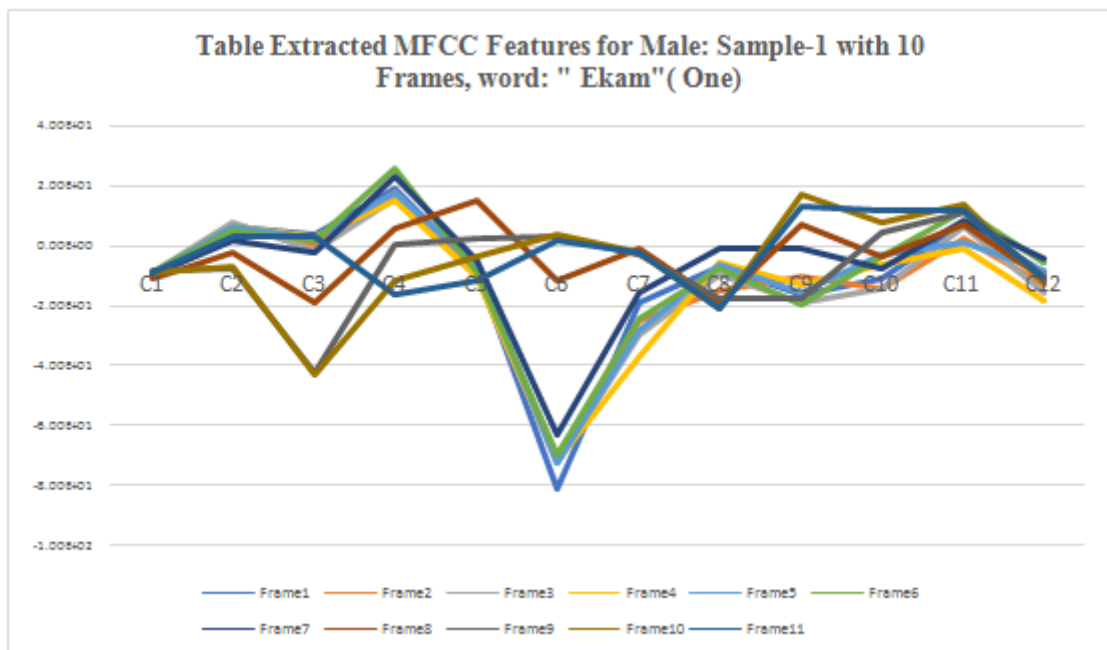


Figure 5: Extracted MFCC Features for Male: Sample word: "Ekam"(One)

VIII. CONCLUSION

To summarize, this study conducted a comparative assessment involving the utilization of Mel Frequency Cepstral Coefficient (MFCC) with 13 coefficients and Support Vector Machine (SVM) for the purpose of recognizing isolated Sanskrit words. Employing a diverse range of techniques and audio characteristics, MFCC exhibited superior accuracy, particularly at the speaker level. The research involved the careful assembly of a Sanskrit speech dataset, featuring a broad spectrum of speakers spanning different age groups. Through an integrated approach, the study achieved an impressive 74% accuracy rate on the training dataset, particularly excelling in the recognition of Sanskrit words and digits. This approach also established predictions based on the correlations within the training data. This work represents a significant contribution to the field of Sanskrit speech recognition, highlighting the potential of MFCC and SVM in enhancing isolated word recognition and speaker differentiation. These findings pave the way for future advancements in human-computer interfaces and speech technology.

REFERENCE

- [1] Pukhraj P. Shrishrimal, Ratnadeep R. Deshmukh, Vishal B. Waghmare "Indian Language Speech Database: A Review", International Journal of Computer Applications (0975 – 888) Volume 47–No.5, June 2012.
- [2] Potale shubham, Kharpude Pratik, Patil rahul, Ajay kumar gupta, "Design and Development of Word Recognition for Marathi Language", Imperial Journal of Interdisciplinary Research (IJIR) Vol-3, Issue- 6, 2017 ISSN: 2454-1362, <http://www.onlinejournal.in>
- [3] Yogesh K. Gedam, Sujata S. Magare, Amrapali C. Dabhade, Ratnadeep R. Deshmukh, "Development of Automatic Speech Recognition of Marathi Numerals - A Review", International Journal of Engineering and Innovative Technology (IJEIT) Volume 3, Issue 9, March 2014.
- [4] Waghmare, Swapnil D. "Stuttered Isolated Spoken Marathi Speech Recognition by using MFCC and LPC." International Journal of Innovations in Engineering and Technology 8.3 (2017).
- [5] Devyani S. Kulkarni, Ratnadeep R. Deshmukh, Pukhraj P. Shrishrimal "A Review of Speech Signal Enhancement Techniques", International Journal of Computer Applications (0975 – 8887) Volume 139 – No.14, April 2016.
- [6] V. K. Kale, R.R. Deshmukh, G. B. Janvale "Development of English Language Isolated Words' Corpus and Views on Recognition Methods", International Journal of Innovative Research in Technology & Science (IJIRTS) · March 2014.
- [7] Ashok Kumar, Vikas Mittal "Speech Recognition: A Complete Perspective", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7 Issue-6C, April 2019
- [8] Yogesh K. Gedam, Paras V. Kothare, Ratnadeep R. Deshmukh "Design and Development of Speech Database of Marathi Numerals", International Journal of Advanced Research in Computer Science and Software Engineering Volume 4, Issue 3, March 2014.
- [9] Suman K. Saksamudre, P.P. Shrishrimal, R.R. Deshmukh, "A Review on Different Approaches for Speech Recognition System", International Journal of Computer Applications (0975 – 8887) Volume 115 – No. 22, April 2015.
- [10] Shaikh Naziya S., R. R. Deshmukh, "LPC and HMM Performance Analysis for Speech Recognition System for Urdu Digits", IOSR Journal of Computer Engineering (IOSR-JCE), Volume 19, Issue 4, Ver. IV. (Jul.-Aug. 2017), PP 14-18.
- [11] Pukhraj P. Shrishrimal, Ratnadeep R. Deshmukh, Vishal B. Waghmare, Sushma Borade, Pooja V. Ganesh B. Janvale "Development of Marathi Language Speech Database from Marathwada Region",

DESIGN AND DEVELOPMENT OF SPEECH DATABASE FOR DIGIT AND VOWELS IN
SANSKRIT LANGUAGE AND RECOGNITION SYSTEM USING MACHINE LEARNING

Conference Paper · October 2015.

- [12] Mr.V.K.Kale,Dr.R.R.Deshmukh,Dr.G.B.Janvale,Mr.V.B.Waghmare,Mr.P.P.Shrishrimal,“Isolated English Words Recognition Spoken by Non-Native Speakers”, International Research Journal of Computer Science (IRJCS) ISSN: 2393-9842 Volume 1 Issue 3 (November 2014).
- [13] P. P. Shrishrimal, R. R. Deshmukh, Vishal B. Waghmare, “development of isolated words speech database of marathi words for agriculture purpose”, Asian Journal of Computer Science and Information Technology 2: 7 (2012) 217 – 218.
- [14] V. B. Waghmare, R. R. Deshmukh, P. P. Shrishrimal, G. B. Janvale “Development of Isolated Marathi Words Emotional Speech Database”, International Journal of Computer Applications (0975 – 8887) Volume 94 – No 4, May 2014.
- [15] Aaron M. Oirere , Ganesh B. JanvaleRatnadeep R. Deshmukh”Automatic Speech Recognition and Verification using LPC, MFCC and SVM” International Journal of Computer Applications (0975 – 8887)
Volume 127 – No.8, October 2015
- [17] Pratik K. Kurzekar, Ratnadeep R. Deshmukh, Vishal B. Waghmare, Pukhraj P. Shrishrimal “A Comparative Study of Feature Extraction Techniques for Speech Recognition System”, International Journal of Innovative Research in Science,Engineering and Technology Vol. 3, Issue 12, December 2014.
- [18] Aaron M. Oirere, Ratnadeep R. Deshmukh, Pukhraj P. Shrishrimal” Development of Isolated Numeric Speech Corpus for Swahili Language for Development of Automatic Speech Recognition System”, International Journal of Computer Applications (0975 – 8887) Volume 74– No.11, July 2013.
- [19] Thiruvengatanadhan, R. "Speech recognition using SVM." International Research Journal of Engineering and Technology (IRJET) 5.9 (2018): 918-921.
- [20] Bai, Jing, Jie Wang, and Xueying Zhang. "A Parameters Optimization Method of v-Support Vector Machine and Its Application in Speech Recognition." J. Comput. 8.1 (2013): 113-120.
- [21] Chavan, Rupali S., and Ganesh S. Sable. "An overview of speech recognition using HMM." International Journal of Computer Science and Mobile Computing 2.6 (2013): 233-238.