Global Review of Business and Technology (GRBT)

Vol. 2, No. 2, July, 2022 ISSN: 2767-1941

AUTOMATIC SPEECH RECOGNITION OF GUJARATI DIGITS USING WAVELET-BASED FEATURES IN GAUSSIAN MIXTURE MODELS AND HIDDEN MARKOV MODELS

Purnima Pandit, The Maharaja Sayajirao University of Baroda, India (purnima.pandit-appmath@msubaroda.ac.in)
Shardav Bhatt, Navrachana University, India (shardavb@gmail.com)

ABSTRACT

In this research, we employ Wavelet-based feature vectors in Hidden Markov Models to achieve automatic recognition of isolated Gujarati speech. To do this, we first created a Gujarati voice dataset of numbers one to ten uttered by eight individuals. The dataset was then enlarged utilizing data augmentation techniques. Wavelet-based feature vectors, Mel-Frequency Discrete Wavelet Coefficients (MFDWC), were then employed for dimensionality reduction and vital feature extraction. These feature vectors were used to compute the observation probabilities using Gaussian Mixture Models (GMM). Further, these probabilities were used in Left-to-right whole-word Hidden Markov Models (HMM) with a different number of states. The Baum-Welch algorithm was used to train the HMM parameters. Finally, the trained HMM was used to classify the test speech. Using confusion matrices and several classification metrics, the classification accuracy of the original and augmented datasets is compared.

Keywords: Automatic Speech Recognition (ASR), Wavelet coefficients, Gaussian Mixture Model (GMM), Hidden Markov Model (HMM), Gujarati language.

1. INTRODUCTION

Automatic Speech Recognition (ASR) is the procedure of creating an intelligent system that can identify a human speech utilizing the data contained in a digital speech signal. Such systems are advantageous for everyday tasks such as hands–free computing, providing instructions to home appliances, automatic voice dialing, and developing a speech user interface. Despite decades of study by researchers, engineers and scientists throughout the world, we are still on a journey of developing an accurate ASR system that can recognize a speech without any error (Juang et al., 2005).

ASR is a multidisciplinary issue (Rabiner et al., 2010). It demands skills and knowledge in a wide range of disciplines including languages, acoustics, signal processing, machine learning, pattern recognition, programming, and applied mathematics and statistics. Another difficulty is that a voice signal contains both background noise and differences in speaker pronunciation.

According to the most recent census statistics (Language: India, State and Union Territories, 2011), there are 55.4 million Gujarati speakers in India. However, due to Gujarat's 79.31 percent literacy rate (Provisional Population totals, 2011), not everyone can use computers, machines or smart gadgets due to a lack of English literacy, particularly in rural regions. ASR is extremely valuable when used to a low-resource Indian regional language such as Gujarati. Such effort becomes important in order to assist individuals who can only communicate in their native tongue Gujarati. People with disability in finger or palm, who are incapable of utilizing input hardware devices, can send commands to computers, machines or smart gadgets using their voice via a speech user interface. Such people would benefit significantly from a speech user interface in the Gujarati language. ASR in the Gujarati language is a challenging problem. Gujarati languages have more phonemes compared to other languages. The language is phonetically distinct from other Indian languages. Correct pronunciation necessitates proper articulations. The vowels એ and ઓ have two different accents (Cardona et al., 2007). Gujarati has a retroflex lateral flap ળ. When compared to other Indian languages, the consonant જ્ઞ is pronounced in a different way.

Feature extraction is the first essential task of ASR. The Mel–Frequency Cepstral Coefficients (MFCC) approach (Davies et al.,1980), which is based on the Fourier transform of a windowed signal, is frequently used for identifying

vital features from the speech data. Instead, we used the wavelet transform in this approach. A wavelet-based feature extraction approach Mel-frequency Discrete Wavelet Coefficients (MFDWC) (Gowdy et al., 2000), was employed in this work. Then for the recognition, Hidden Markov Models (HMM) and Gaussian Mixture Models (GMM) are used jointly to classify speech characteristics.

The incorporation of Wavelet-based features at the front end is a novel aspect of this study. Wavelets are helpful for methodically removing noise from an audio signal, resulting in a more accurate representation of the information than the original signal. Furthermore, as compared to other de-noising techniques, the calculations of the wavelet coefficients are quicker (Shukla et al., 2013).

The paper is divided into six sections. It begins with an introduction, followed by a review of the literature and a discussion of the objectives. The methodologies are then briefly outlined in section 4. The fifth part goes over the experiments and outcomes in depth. The paper concludes with sections of conclusions and suggestions for further research.

2. LITERATURE REVIEW

In the recent decade, researchers have begun to investigate the Gujarati language for speech recognition studies. Progress has been assessed by a few authors. The authors of (Kurian, 2014) examined this development for various languages, including Gujarati. The most recent survey is available in (Parikh et al., 2020). According to these studies, Gujarati is a low-resource language, and further study is needed for it. Many researchers are working in the generation of Gujarati speech data. The writers of (Malde et al., 2013) and (Madhavi et al., 2014) generated speech data for Indian languages, including Gujarati. The data for emotion detection is created over the course of the work of (Tank et al., 2020). Researchers are employing a variety of strategies for recognition, including Vector Quantization (Chauhan et al., 2015), Artificial Neural Networks (Patel et al., 2013) (Desai et al., 2016), and Support Vector Machines (Chittora et al., 2014). The HMM technique is employed in the works of (Patel et al., 2015), (Valaki et al., 2017), and (Tailor et al., 2018). Some of the most recent works are based on Machine Learning and Deep Learning (Raval et al., 2022). In contrast to all the previous studies, we used a wavelet-based speech features MFDWC in conjunction with a GMM-based HMM in our study. As a result of this literature review, we outline the objectives of our study, which distinguishes our work from existing publications.

3. OBJECTIVES

Previously, we employed Dynamic Time Warping (Pandit et al., 2014), Artificial Neural Networks (Pandit et al., 2014a), Radial Basis Function Network (Pandit et al., 2016) for the recognition, and MFCC for feature extraction. We changed the feature extraction approach to MFDWC in (Pandit et al., 2017) and in (Pandit et al., 2021). In this paper, we aim to evaluate the performance of HMM in categorizing digits 1 to 10 spoken in Gujarati using wavelet-based features. Our objectives are essentially divided into five stages, as shown in Fig. 1.

Figure 1: Objectives of this work

The first and most important objective is to generate speech data in form of recordings of digits one to ten pronounced in Gujarati by diverse speakers. The data is then pre-processed by removing noise and cropping the silent zone. The most important objective of this study is feature extraction, which involves finding important features from speech using wavelet coefficients. The next objective is to generate HMM for each word for all the speakers. The acoustic model of the HMM can be constructed in a variety of ways, but our objective is to employ GMM. Finally, the classification step's goal is to classify the unknown test word in one of the ten classes defined by the ten HMMs.

4. METHODOLOGIES

The approaches utilized for our voice recognition experiments are divided into two stages. The first stage is the feature extraction process using MFDWC. Then, in the second stage, an acoustic model is built using GMM employing these features. HMM incorporates the observation probabilities obtained from the acoustic model. The features of unknown speech are matched to HMMs built for each word. The model with the highest score is chosen as the best model.

4.1 Feature extraction

Most commonly used methods for feature extraction are Linear Predictive Coding, Perceptual Linear Prediction, Mel Frequency Cepstral Coefficients (Juang et al., 2008). We have used a wavelet-based feature extraction method: Mel-Frequency Discrete Wavelet Coefficients (MFDWC) (Gowdy et al., 2000). To determine these features, first step is pre-emphasizing the signal. In this step, first order filter defined by equation (1), is applied on the speech signal $x(n)$ to make signal less prone to noise.

$$s(n) = x(n) - \alpha\, x(n-1) \tag{1}$$

Here, n is a sample number and α is constant. Since the speech signal is dynamically varying signal, the next step is to divide the signal into number of frames with frame size N and having M overlapping samples with the neighboring frame. So, we get

$$s_i(t), \qquad 1 \le t \le N \tag{2}$$

Here, i is the frame number. This step is required so that in a small frame the variations are less. To minimize the discontinuities in the signal, the Hamming window (3) is applied on each frame s_i.

$$w(t) = (1-\beta) - \beta \cos\left(\frac{2\pi t}{N-1}\right) \tag{3}$$

In the next step, the energy of the absolute value of Fourier transform, i.e. power spectrum, is determined for each frame using (4).

$$p_i(k) = \frac{1}{N}\left|\sum_{t=0}^{N-1} s_i(t)\, w(t)\, e^{-\frac{2j\pi kt}{N}}\right|^2 \tag{4}$$

Here, $1 \le k \le N/2$. This step is necessary, to do the analysis in the frequency domain. Now these frequencies need to be converted to mels using (5) because human ears receive frequency on the logarithmic scale.

$$M(f) = 1125 \ln\left(1 + \frac{f}{700}\right) \tag{5}$$

After shifting to the mel-scale, a filter bank of triangular filters (6) is applied on (4) for each frame, which gives (7).

$$F_j(k) = \begin{cases} \dfrac{k - M(j-1)}{M(j) - M(j-1)}, & M(j-1) \le k \le M(j) \\ \dfrac{M(j+1) - k}{M(j+1) - M(j)}, & M(j) \le k \le M(j+1) \\ 0, & \text{Otherwise} \end{cases} \tag{6}$$

$$\tilde{p}_m = \sum_{k=0}^{N/2} p_i(k)\, F_j(k) \tag{7}$$

Here, j is a filter index, $M(j)$ are mel-scaled frequencies. Finally, the MFDWC feature vectors are calculated by applying the Discrete Wavelet Transform to the mel-scaled log filter bank energies of each speech frame.

The general expression of a discrete wavelet transforms (DWT) for a discrete signal $X[n]$, having M samples, is given by approximations (8) and details (9).

$$W_\phi[j_0, k] = \frac{1}{\sqrt{m}} \sum_n X[n] \phi_{j_0,k}[n] \tag{8}$$

$$W_\psi[j, k] = \frac{1}{\sqrt{m}} \sum_n X[n] \psi_{j,k}[n], \quad j \ge j_0 \tag{9}$$

Here $\phi_{j_0,k}$ is a discrete scaling function and $\psi_{j,k}[n]$ is a discrete wavelet function having M components each (Liu, 2010). Approximations and details give the DWT of a given signal. Applying the DWT on (7), we obtained 13 coefficients per frame for each signal. These coefficients are used to train HMM parameters for each word.

4.2 Hidden Markov Model (HMM)

The HMM comprises of N states $\{S_1, S_2, \dots, S_N\}$ and T observations $\{o_1, o_2, \dots, o_T\}$. The states are hidden. The parameters of HMM are initial state probabilities $\pi_i = P(S_i \text{ at time } t = 1)$, state transition probabilities $a_{ij} = P(S_j \text{ at time } t+1 \mid S_i \text{ at time } t)$ and observation probabilities $b_j(o_k) = P(o_k \text{ at time } t \mid S_j \text{ at time } t+1)$. These HMM parameters are trained using Baum-Welch algorithm (Rabiner et al., 1989). Let $\xi_t(i,j)$ = Joint probability of being in state S_i at time t and state S_j at time $t+1$ and $\gamma_t(i)$ = Probability of being in state S_i at time t. The algorithm is given by equations (10)-(12).

$$\pi^* = \gamma_1(i) \tag{10}$$

$$a_{ij}^* = \sum_{t=1}^{T-1} \xi_t(i,j) \Big/ \sum_{t=1}^{T-1} \gamma_t(i) \tag{11}$$

$$b_j^*(o_k) = \sum_{\substack{t=1 \\ \text{s.t. } o_t = S_k}}^{T-1} \gamma_t(j) \Big/ \sum_{t=1}^{T-1} \gamma_t(j) \tag{12}$$

The HMM with trained probabilities are then used for testing of unknown observations.

4.3 Gaussian Mixture Model (GMM)

For HMM, we need to determine observation likelihood $b_j(o_k)$ i.e., $P(o_k \text{ at time } t \mid S_j \text{ at time } t+1)$. Here the observations are real-valued feature vectors. We have to represent it by a probability density function (PDF). GMM uses a linear combination of multivariate Gaussian distributions to determine the observation likelihood $b_j(o_k)$ using equation (13) (Jurafsky et al., 2008).

$$b_j(o_k) = \sum_{m=1}^{M} c_{jm} \frac{1}{\sqrt{2\pi|\Sigma_{jm}|}} e^{-\frac{1}{2}(o_k - \mu_{jm})^T \Sigma_{jm}^{-1} (o_k - \mu_{jm})} \tag{13}$$

Here Σ_{jm} is the covariance matrix and μ_{jm} is the mean for m^{th} Gaussian PDF and for the j^{th} feature vector. c_{jm} are the coefficients of GMM. Probabilities calculated this way are useful in HMM for the classification of words.
The approaches described here are summarized in the Fig. 2, in the order indicated. The next section explains the experiments based on these methodologies.

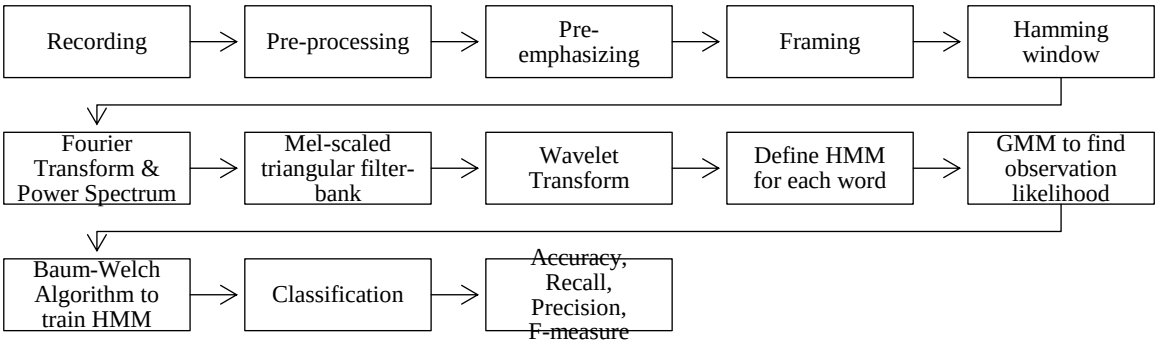

Figure 2: Methodologies used in this work

5. EXPERIMENTS AND RESULTS

All the experiments are done using an open-source programming language Python. The experiments comprise of classification of Gujarati digits using MFDWC and HMM. For this, two python functions were coded. Standard libraries like NumPy, SciPy, Matplotlib, Pandas, PyWavelets, Scikit-Learn were included in the code.

5.1 Data generation and pre-processing

The audio data was generated by recording the speech in the 32-bit PCM wav format with the sampling rate of 16,000 samples per seconds. The recording was done using common headphones in a normal room environment. The open-source software Audacity was used for the pre-processing. The pre-processing consists of cropping unvoiced part from speech and removing noise from the voiced part. The recording consists of digits one to ten uttered in Gujarati language by eight speakers. This makes the database of 80 wav files. Data augmentation techniques like amplifying/de-amplifying the signal were used to expand the dataset. Moreover, a random noise was also added, to make it more realistic for the applications. This way the expanded dataset consists of 160 samples.

5.2 Feature Extraction

For the feature extraction, MFDWC method was used. In the pre-emphasis step, $\alpha = 0.97$ in equation (1) was considered. Each word was divided into frames having length $N = 256$ with $M = 100$ samples overlapping. The hamming with $\beta = 0.46$ in equation (3) was applied on each frame. The filter-bank consists of 20 triangular filters. Fig. 3 to Fig. 6 explains how MFDWC feature vectors of length 13 were determined using the filters associated with Daubechies – 6 wavelets as shown in the Fig. 7.

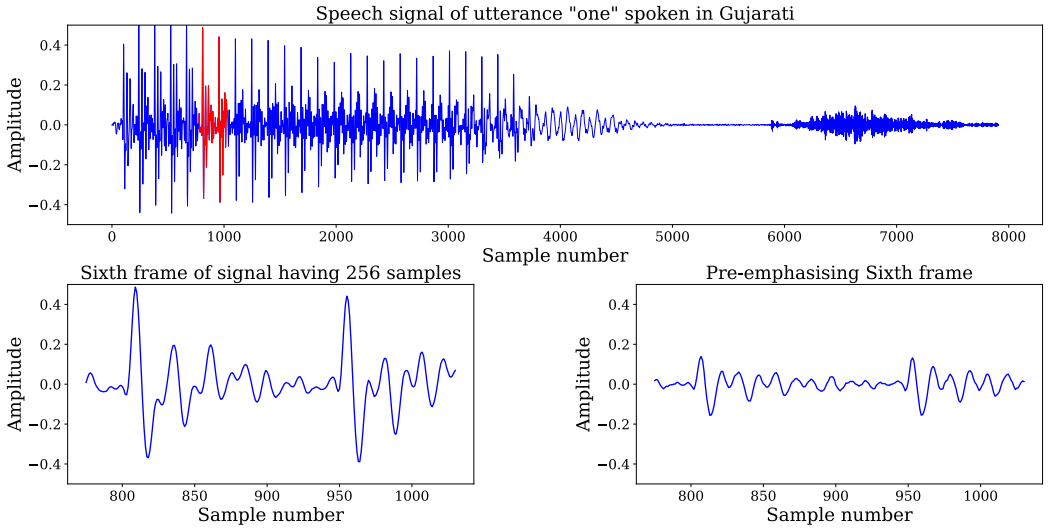

Figure 3: Speech Signal and its framing and pre-emphasizing

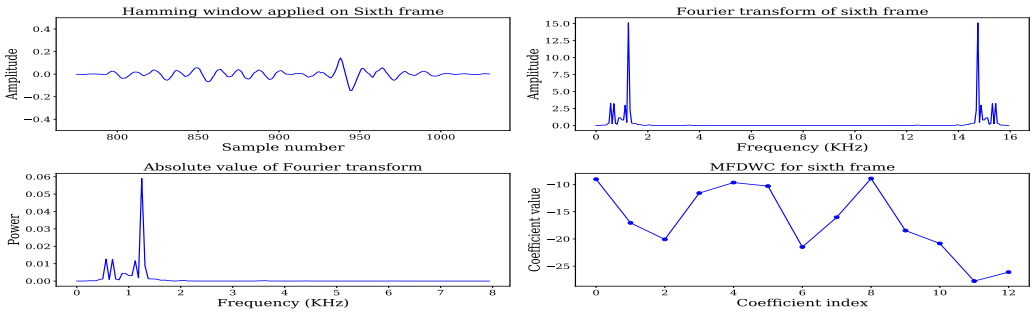

Figure 4: Various steps of MFDWC applied on sixth frame of the signal

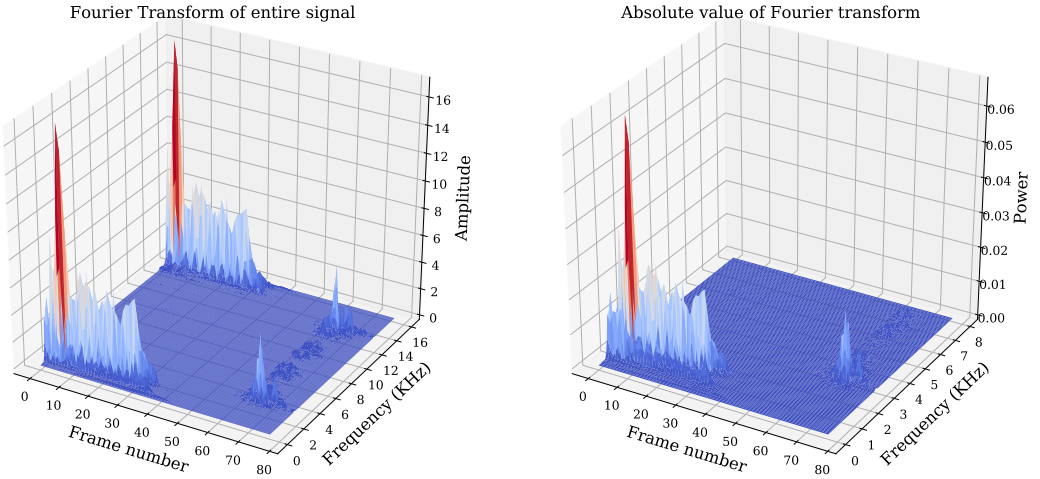

Figure 5: Fourier transform and its power spectrum of entire signal

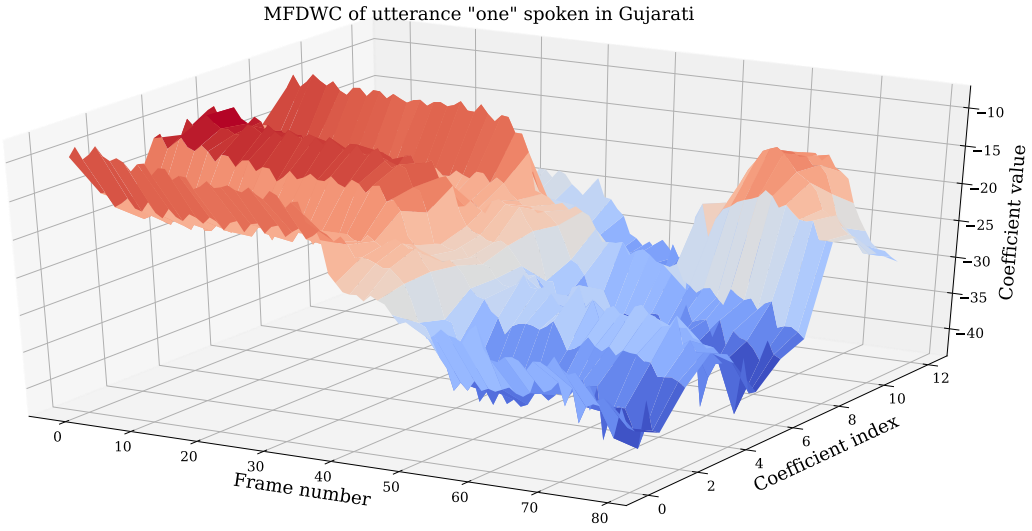

Figure 6: MFDWC of entire signal

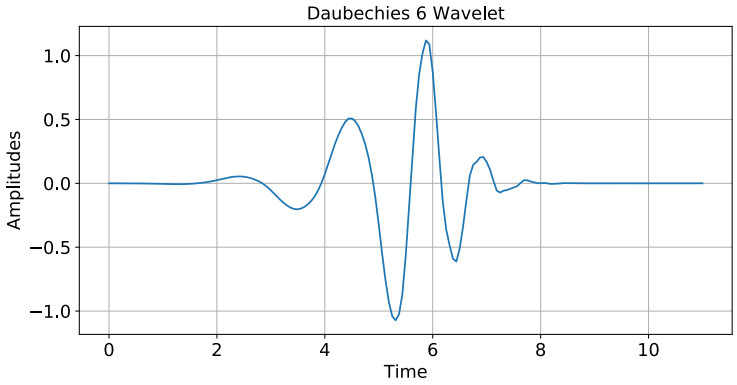

Figure 7: Daubechies-6 wavelet function

5.3 Recognition

For the recognition part, the HMM with GMM was used. The left-to-right HMM was created for each word. In these HMMs, the hidden states were the speech units hidden inside the recording and the observations were the MFDWC feature vectors. Different number of states from 2 to 7 were considered and the accuracy was determined in each case. The HMM parameters π_i and a_{ij} were initialized randomly while the parameter $b_j(o_k)$ was initiated by determining the probability distribution of MFDWC features using GMM defined by (13). In this, the mean and the variance of each feature vector for each dimension was determined. The dataset was divided into training and testing using stratified sampling method with proportion of 12.5% of test patterns. Then these parameters were trained using equations (10)-(12) of Baum-Welch algorithm. During testing, given an unknown word, these trained parameters are useful to determine the HMM of the unknown word and this way we can identify the spoken word.

5.4 Results

The results obtained in various experiments are summarized in the Table 1. It displays the training time for the HMM in seconds as well as the accuracy gained when testing the unknown speech. These two parameters were calculated in the HMM for various number states for both datasets: original and augmented.

Table 1: Summary of experiments performed

	Original Dataset		Augmented Dataset	
Number of States in HMM	Training Time (sec)	Test Accuracy (%)	Training Time (sec)	Test Accuracy (%)
2	5.17	100	14.14	60
3	5.38	100	14.93	65
4	5.62	100	16.33	65
5	5.76	100	16.15	70
6	6.19	100	17.14	65
7	6.33	100	17.78	65

The test accuracy was determined using,

$$\text{Test accuracy} = \frac{\text{Correctly classified test patterns}}{\text{Total number of test patterns}} \times 100 \tag{14}$$

Fig. 8 and Fig. 9 shows confusion matrices between the test data and the predicted data for the original dataset and the augmented dataset. These confusion matrices are prepared for the 5-state HMMs for both Original and the augmented dataset. The correct classification yields the entry on the diagonals of the matrix.

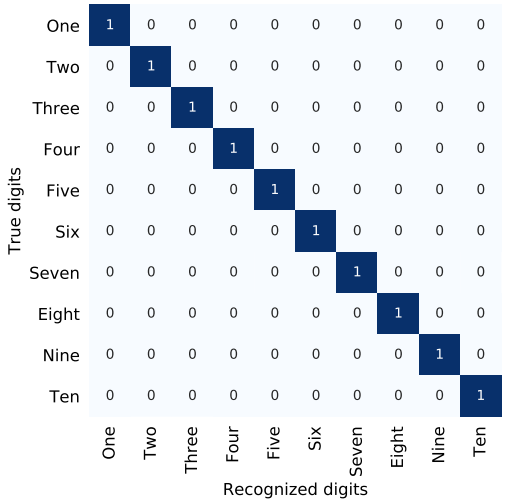

Figure 8: Confusion matrix for the original dataset

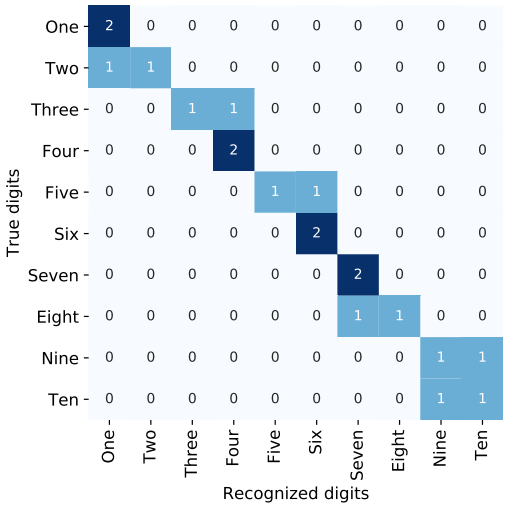

Figure 9: Confusion matrix for the augmented dataset

Various classification measures like Recall, Precision and F-measure are calculated using (15)-(17) (Fawcett, 2006) for the 5-state HMMs for both the original and the augmented dataset to get more insight into how well the classification is done.

$$\text{Recall} = \frac{\text{True postive}}{\text{Positive}} \tag{15}$$

$$\text{Precision} = \frac{\text{True postive}}{\text{True postive} + \text{False positive}} \tag{16}$$

$$\text{F} - \text{measure} = \frac{2}{\frac{1}{\text{Recall}} + \frac{1}{\text{Precision}}} \tag{17}$$

The values of these classification measures are summarized in Table 2.

Table 2: Summary of various classification measures

	Original Dataset			Augmented Dataset		
Digits	Recall	Precision	F- measure	Recall	Precision	F- measure
1	1.00	1.00	1.00	0.67	1.00	0.80
2	1.00	1.00	1.00	1.00	0.50	0.67
3	1.00	1.00	1.00	1.00	0.50	0.67
4	1.00	1.00	1.00	0.67	1.00	0.80
5	1.00	1.00	1.00	1.00	0.50	0.67
6	1.00	1.00	1.00	0.67	1.00	0.80
7	1.00	1.00	1.00	0.67	1.00	0.80
8	1.00	1.00	1.00	1.00	0.50	0.67
9	1.00	1.00	1.00	0.50	0.50	0.50
10	1.00	1.00	1.00	0.50	0.50	0.50

6. CONCLUSION AND FUTURE WORK

6.1 Conclusion

We used HMM to conduct ASR studies on isolated words uttered in the Gujarati language. The wavelet-based coefficients, MFDWC, were employed to collect vital features from the recorded speech. The word HMM was developed to classify the digits. The GMM was utilized to calculate the HMM's observation probability. These trials yielded 100 % accuracy for the original dataset. Moreover, 70 % accuracy was achieved for the augmented dataset with noise, which is a promising result.

6.2 Future work

Because these were our initial HMM studies, the dataset was limited and only consisted of words. As a result, one of the focuses of our future work will be to expand the dataset and add continuous phrases into it. In order to make this research more applicable for the applications, advanced signal processing algorithms for extracting words from sentences must be researched. Instead of the GMM technique, some researchers have investigated a hybrid approach that combines HMM with Artificial Neural Networks. In the future, we aim to incorporate Deep Learning and Convolutional Neural Networks into our experiments of Speech recognition for the Gujarati language.

REFERENCES

Cardona, G., & Jain, D. (2007). The Indo-Aryan Languages. Routledge.

Chauhan, H. B., & Tanawala, B. A. (2015). Comparative Study of MFCC And LPC Algorithms for Gujrati Isolated Word Recognition. *International Journal of Innovative Research in Computer and Communication Engineering*, 3(2), 822–826. https://doi.org/10.15680/ijircce.2015.0302056

Chittora, A., & Patil, H. A. (2014). Classification of phonemes using modulation spectrogram based features for Gujarati language. *International Conference on Asian Language Processing (IALP*), 46–49. https://doi.org/10.1109/IALP.2014.6973506

Davis, S. B., & Mermelstein, P. (1980). Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(4), 357–366. https://doi.org/10.1109/TASSP.1980.1163420

Desai, V. A., & Thakar, V. K. (2016). Neural Network Based Gujarati Speech Recognition for Dataset Collected by in-ear Microphone. *Procedia Computer Science*, 93, 668–675. https://doi.org/10.1016/j.procs.2016.07.259

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Juang, B. H., & Rabiner, L. R. (2005). Automatic Speech Recognition – A Brief History of the Technology. 1–24.

Jurafsky, D., & Martin, J. H. (2008). Speech and Language Processing: An Introduction to Natural Language Processing, *Computational Linguistics, and Speech Recognition* (2nd ed.). Prentice Hall.

Kurian, C. (2014). A Survey on Speech Recognition in Indian Languages. *International Journal of Computer Science and Information Technologies*, 5(5), 6169–6175.

Language: India, States and Union Territories (Table C-16). (2011). In Census of India 2011. Office of The Registrar General, India. http://censusindia.gov.in/2011Census/C-16_25062018_NEW.pdf. Accessed 15 March 2022.

Liu, C.-L. (2010). A Tutorial of the Wavelet Transform.

Madhavi, M. C., Sharma, S., & Patil, H. A. (2014). Development of language resources for speech application in Gujarati and Marathi. *International Conference on Asian Language Processing (IALP),* 115–118. https://doi.org/10.1109/IALP.2014.6973517

Malde, K. D., Vachhani, B. B., Madhavi, M. C., Chhayani, N. H., & Patil, H. A. (2013). Development of speech corpora in Gujarati and Marathi for phonetic transcription. *International Conference Oriental COCOSDA Held Jointly with 2013 Conference on Asian Spoken Language Research and Evaluation*, O-COCOSDA/CASLRE 2013, 1–6. https://doi.org/10.1109/ICSDA.2013.6709865

Pandit, P., & Bhatt, S. (2014). Automatic Speech Recognition of Gujarati digits using Dynamic Time Warping. *International Journal of Engineering and Innovative Technology*, 3(12), 69–73.

Pandit, P., Bhatt, S., & Makwana, P. (2014). Automatic Speech Recognition of Gujarati Digits using Artificial Neural Network. *Proceedings of 19th Annual Cum 4th International Conference of GAMS on Advances in Mathematical Modelling to Real World Problems*, 141–146.

Pandit, P., & Bhatt, S. (2016). Automatic Speech Recognition of Gujarati digits using Radial Basis Function Network. *Proceedings of International Conference on Futuristic Trends in Engineering, Science, Pharmacy and Management*, 216–226.

Pandit, P., & Bhatt, S. (2017). Automatic Speech Recognition of Gujarati digits using Wavelet Coefficients. *Journal of The Maharaja Sayajirao University of Baroda*, 52(1), 101–110.

Pandit, P., Makwana, P., & Bhatt, S. (2021). Automatic Speech Recognition of Continuous Speech Signal of Gujarati Language Using Machine Learning. Mathematical Modeling, *Computational Intelligence Techniques and Renewable Energy*, 147–159. https://doi.org/https://doi.org/10.1007/978-981-15-9953-8_13.

Parikh, R. B., & Joshi, H. (2020). Gujarati Speech Recognition – A Review. Test Engineering and Management, 83, 549–553.

Patel, J., & Nandurbarkar, A. (2015). Development and Implementation of Algorithm for Speaker recognition for Gujarati Language. *International Research Journal of Engineering and Technology*, 2(2), 444–448.

Patel, P., & Jethva, H. (2013). Neural Network Based Gujarati Language Speech Recognition. *International Journal of Computer Science and Management Research*, 2(5), 2623–2627.

Rabiner, L. R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, 77(2), 257–286.

Rabiner, L. R., Juang, B.-H., & Yegnanarayana, B. (2010). Fundamentals of Speech Recognition (2nd ed.). *Pearson Education*.

Raval, D., Pathak, V., Patel, M., & Bhatt, B. (2022). Improving Deep Learning based Automatic Speech Recognition for Gujarati. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 21(3), 1–18. https://doi.org/10.1145/3483446

Registrar General and Census Commissioner. (2011). Provisional population totals. In Census of India 2011 (Vol. 1). Office of The Registrar General and Census Commissioner, India. https://censusindia.gov.in/2011-prov-results/prov_results_paper1_india.html. Accessed 14 March 2022.

Shukla, K. K., & Tiwari, A. K. (2013). Efficient Algorithms for Discrete Wavelet Transform. *Springer*, London. https://doi.org/10.1007/978-1-4471-4941-5

Tailor, J. H., & Shah, D. B. (2018). HMM-Based Lightweight Speech Recognition System for Gujarati Language. *Lecture Notes in Networks and Systems*, 10, 451–461. https://doi.org/10.1007/978-981-10-3920-1_46

Tank, V. P., & Hadia, S. K. (2020). Creation of speech corpus for emotion analysis in Gujarati language and its evaluation by various speech parameters. *International Journal of Electrical and Computer Engineering*, 10(5), 4752–4758. https://doi.org/10.11591/ijece.v10i5.pp4752-4758

Tufekci, Z., & Gowdy, J. N. (2000). Feature extraction using discrete wavelet transform for speech recognition. *Conference Proceedings - IEEE SOUTHEASTCON*, 116–123.

Valaki, S., & Jethva, H. (2017). A hybrid HMM/ANN approach for automatic Gujarati speech recognition. International Conference on Innovations in Information, *Embedded and Communication Systems (ICIIECS)*, 1–5. https://doi.org/10.1109/ICIIECS.2017.8276141.