Tribal Language Identifier

Reshma Raj[1], Riyas K. S[2], Dr. Jayan A. R[3]
[1, 2, 3] Department of Electronics and Communication Engineering
[1] Government Engineering College wayanad
[2, 3] RIT, Kottayam

Abstract- *Automatic language identification involves the decision of the language in which speaker being spoken. In this paper, we describe a method to perform language identification analysis for 2 languages, to be specific Paniya and Malayalam. Here, we take the advantages of bessel properties as another option to the prevalent procedures like MFCC and LPCC. The pseudo- stationary signals such as speech signal can be effectively represented by using damped sinusoidal bessel basis functions. So, a periodic and damped sinusoidal signal like bessel function are best for experimental analysis of speech signal. A set of 12 speakers of each of 2 languages are used. The suggested system is tested over the database containing Malayalam and Paniya and acquired an efficiency of 99%.*

Keywords- FBCC, Feature Extraction, Language Identification, MFSC, Mel Filter

I. INTRODUCTION

Speech serves to convey data from a speaker to one or more audience members. It results from a combination of sound energy modulated by a filter transfer function dictated by the supra laryngeal vocal tract. It conveys important linguistic information in communication among human beings. The speech signals encompass numerous information. Basically a message is passed on by means of talked words and it contains information about the feeling, character of speaker, gender and language being talked.

An automatic language identification system use speech messages as input and produce the identity of the language being talked as the output. Malayalam is a Dravidian language fundamentally talked in the southwest of India. The Paniya, otherwise called Paniyar or Paniyan, are also a member of Dravidian family of India. They primarily inhabit in Kerala. Specifically in the Wayanad and the neighboring parts of Kannur, Kozhikode and Malappuram districts. The Paniya talk the Paniya language as a native language. It is most firmly related to Malayalam, Kadar, Ravula and other Malayalam languages. This paper involves the design of an efficient language identification system for recognizing Malayalam and Paniya. An automatic language identification task comprised of mainly 2 phases:

1) Feature Extraction &
2) Classification

In feature extraction stage, the desired features are extracted from the speech samples thus form feature database. In classification there are 2 steps, to be specific training and recognition (also called testing). During training, the speech samples of languages to be identified are examined and produce models for each of languages. These language models describe characteristics of the training samples and the dependency among language and then these models can be utilized during second stage of identification system namely, recognition phase. During testing, formerlyinconspicuous test sample is applied to the system and yield the language that most nearly matches the test message.

II. PROPOSED METHOD

Figure. 1 given below depicts the methodology of the proposed work. In first phase, the features from speaker's database are collected and stored it in a feature database is now accessed by the classifier to recognize new data. In the language identification, the language whose model best matches with the test utterance is declared as the identified language.

2.1 Speech Database Collection

The database for this proposed work comprised of 20 talked words from 6 male and 6 female speakers of each of 2 languages namely, Malayalam and Paniya.

2.2 Feature Extraction

In feature extraction phase, we extract the features of speech samples and put away it in a feature database. i.e., the feature extraction transform the speech signal into an another form of signal or a set of signal or set of parameters with an objective of simplify the speech signal or to remove the redundancy present in the speech signal.

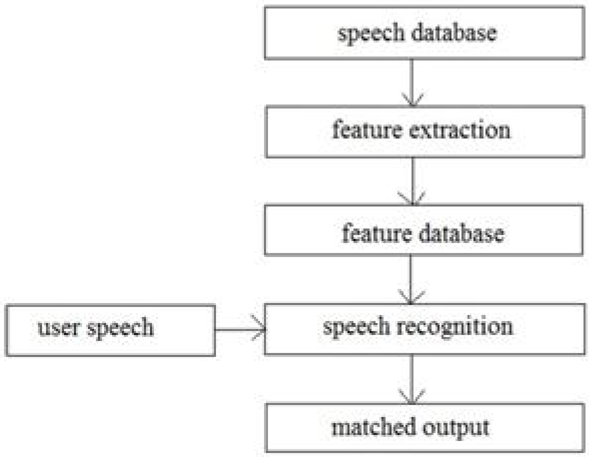


Figure 1: Proposed Methodology

In this proposed work, we extract 3 features:

1) Fourier Bessel Cepstral Coefficient (FBCC)
2) Parseval Energy &
3) Signal Energy

2.2.1 Fourier Bessel Cepstral Coefficients

The block diagram given below depicts the steps for the extraction of FBCC from a speech signal. The preprocessing stage is utilized as part of request to increase the efficiency of subsequent feature extraction and classification stages and thus to enhance the overall system performance. In speech analysis, it is assumed that the speech signal properties change slowly within a short time window. i.e., the extracted features presumed to remain fixed for a short time window. So we must divide the speech signal into successive frames or windows.

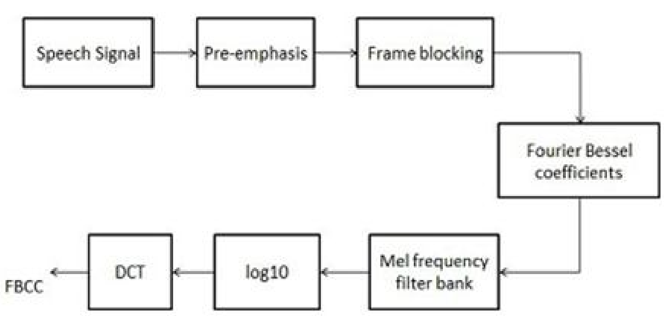


Figure 2: Block diagram of extraction of FBCC

Sinusoidal basis function cannot make an efficient representation of irregular non stationary speech signal. Aperiodic and damped sinusoidal such as bessel function of zero order are best suited for analysis of speech signal [1]. For discrete time signal, y[n] which is defined over the interval [0, N], the zero order Fourier Bessel series can be written as,

$$y[n] = \sum_{m=1}^{M} Cm\, J0\left(\lambda m \frac{n}{N}\right)$$

Where,

J_0 () - Bessel function of zero order

$\{\lambda m : m = 1, \ldots M\}$ - ascending order positive roots of J0 ()=0

Cm - Fourier Bessel Coefficients

The Fourier Bessel Coefficients,

$$Cm = \frac{2\sum_{n=1}^{N} n\, y[n]\, J0\left(\lambda m \frac{n}{N}\right)}{N^2 J_1(\lambda m)^2}$$

Where,

J1 ()- Bessel function of first order

To improve the performance, the Fourier Bessel coefficients are applied to a Mel filter [9]. Output of each filter in a Mel filter bank corresponds to the total energy in frequencies that lie within the range of that filter. Then, calculate the logarithm of filter energies, it results Mel Frequency Spectral Coefficients (MFSC). DCT of this Mel Frequency Spectral Coefficients are called Fourier Bessel Cepstral Coefficients (FBCC).

2.2.2. Parseval Energy

Each Fourier Bessel coefficient has corresponding instantaneous parseval energy,

$$E_m = {C_m}^2 \frac{N^2}{2} J_1(\lambda_m)^2$$

i.e., there is a one to one correspondence between the paresval energy and Fourier Bessel coefficient [1]. Block diagram given below shows the steps for extracting Parseval energy from a speech sample.

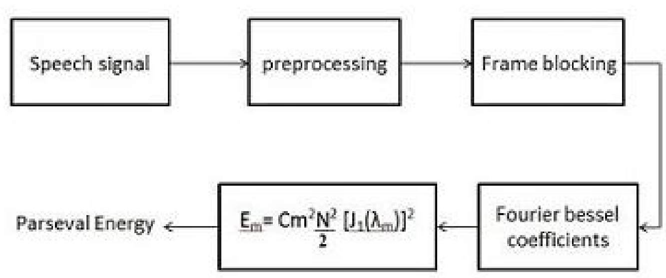


Figure 3: Block diagram for the extraction of parseval Energy

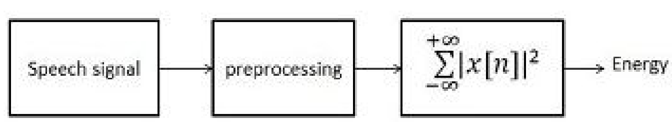


Figure 4: Block diagram for the extraction of Signal Energy

2.2.3. Signal Energy

The term signal energy is used to characterize a signal. For a discrete time signal y[n] defined over an arbitrary interval [0, N], signal energy is,

$$E_y = \sum_{n=-\infty}^{\infty} |y[n]|^2$$

Block diagram for the extraction of the signal is energy is shown in Figure. 4.

III.CLASSIFICATION

People are often make mistakes during experimental analysis or, possibly, when trying to find relationships between multiple features. This makes it tough for them to find results to some problems. In such situations, machine learning can often successfully apply to increase the efficiency of systems and the designs of machines. In this proposed work, we use 2 Classifiers.

1) Support vector Machine (SVM) &
2) Random Forest

IV. RESULTS

In this method we first preprocess the speech samples to make the feature extraction more effective. Figure (a) shows a preprocessed speech sample.

Then performed the Feature extraction stage using MATLAB software. In this stage, we extract the features namely Fourier Bessel Cepstral Coefficients (FBCC), Parseval energy and Signal energy from the speech samples. Once we find Fourier Bessel Coefficients of a speech sample, we can reconstructthe speech signal by using that resultant Fourier Bessel Coefficients. Such a reconstructed speech signal and the bessel function of zero order is shown in Figure (b) and (c) respectively.

After feature extraction, the resultant feature vectors are stored in a feature database and then it is used as input vector in the classification stage. Classification is performed in WEKA platform. After discretizing and standardizing, the data is evaluated with test option of percentage split 10%, which means that 10% data go for training and 90% for testing.

Tables 1 and 3 shows the confusion matrix of SVM and Random Classifiers. Here there are 2 possible classes, namely Malayalam and Paniya. Classifier made a total of 2160 predictions. Out of 2160 predictions, SVM classifier predicts 'Malayalam' 1075 times and out of these 1075 predictions, 4 predictions are incorrect. Similarly, the classifier made a total of 1085 times predictions for 'Paniya' and out of these 1085 predictions, 88 predictions are incorrect. So the classifier gave an accuracy rate,

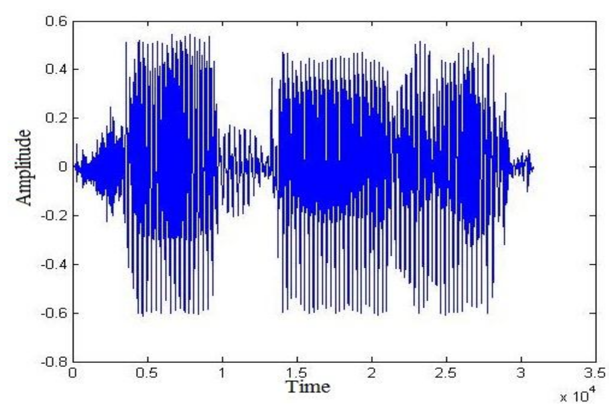
$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$
$$= \frac{1071 + 1077}{1071 + 4 + 1077 + 88} = 99.4444\,\%$$

For Random Forest, Out of 2160 predictions, the classifier predicts 'Malayalam'1084 times and out of this 1084 predictions, 33 predictions are incorrect. Similarly, the classifier made a total of 1076 times predictions for 'Paniya' and out of these 1076 predictions, 28 predictions are incorrect. So the classifier gave an accuracy rate,

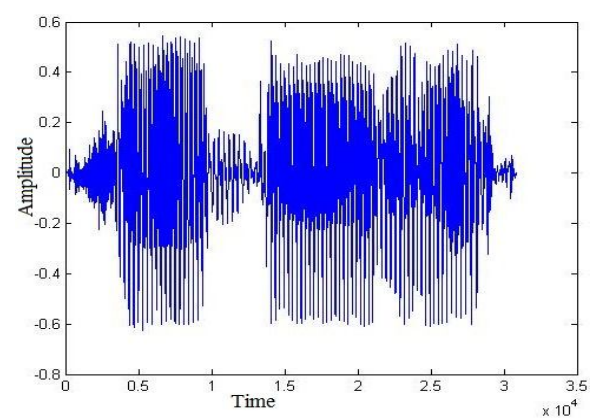
$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$
$$= \frac{1051 + 1048}{1051 + 33 + 1048 + 28} = 97.1757\,\%$$

So the different classifier gives different results. Here, SVM classifier provides high accuracy rate than Random Forest. i.e., SVM is best suit for this proposed work compared to Random Forest.

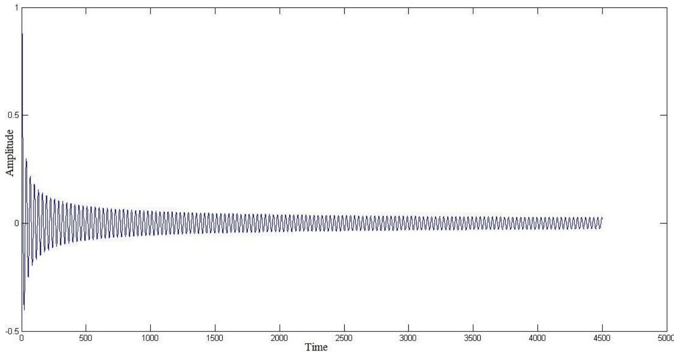
V. FIGURES AND TABLES



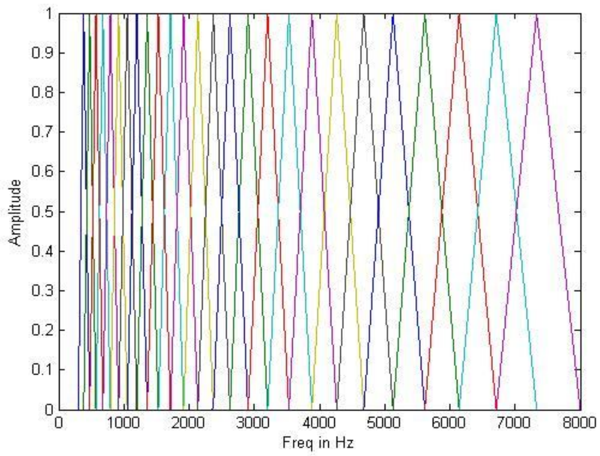
(a) A speech .wav of an utterance "Chithalu"



(b) Reconstructed .wavof an utterance “Chithalu”



(c) Bessel function of zero order



(d) A Mel filter bank consisting 10 filters

Table 1: Confusion matrix of SVM classifier

No. of instances: 2160	Predicted Malayalam	Predicted Paniya	
Actual Malayalam	1071	8	1079
Actual Paniya	4	1077	1081
	1075	1085	

Table 2: System performance measurements of SVM

	Malayalam	Paniya
Accuracy	99.4444 %	
Mis classification rate	0.5556 %	
TP rate	0.996	0.993
FP rate	0.007	0.004
Precision	0.993	0.996

Table 3: Confusion matrix of Random Forest classifier

No. of instances: 2160	Predicted Malayalam	Predicted Paniya	
Actual Malayalam	1051	28	1079
Actual Paniya	33	1048	1081
	1084	1076	

Table 4: system performance measurements of Random Forest

	Malayalam	Paniya
Accuracy	97.1757 %	
Mis classification rate	2.8241 %	
TP rate	0.969	0.974
FP rate	0.026	0.031
Precision	0.974	0.970

VI. CONCLUSION

In this proposed work, the system could achieve a satisfactory performance with the features namely Fourier Bessel Cepstral Coefficients (FBCC), Paresval Energy and Signal Energy. With SVM classifier, system achieved an 99.44% correct identification. For Random Forest classifier, the system accuracy is 97.18%. So we can say that Support Vector Machine (SVM) is more suitable for this proposed work.

REFERENCES

[1] K. Gopalan, Timothy R. Anderson and Edward J. Cupples. A Comparisonof Speaker Identification Results Using Features Based on Cepstrum andFourier Bessel Expansion. IEEE Transactions On Speech AndAudio Processing, Vol. 7, No. 3,(May 1999).

[2] K.P. Li and T. J. Edwards, “Statistical models for automatic languageidentification”, In Proceedings 1980 IEEE International Conference on Acoustics, Speech, and Signal Processing, Denver, CO, April 1980.

[3] F.J. Goodman, A.F. Martin, and R.E. Wohlford“Improved automatic language identification in noisy speech”, In Proceedings 1989 IEEE International Conference on Acoustics, Speech, and Signal Processing, Glasgow, Scotland, May 1989

[4] D. Cimarusti and R. B. Ives, “Development of an automatic identificationsystem of spoken languages: Phase 1”, In Proceedings 1982 IEEE International Conference on Acoustics, Speech, and Signal Processing, Paris, France, May 1982

[5] J. T. Foil “Language identification using noisy speech”, In Proceedings 1986 IEEE International Conference on Acoustics, Speech, and Signal Processing, Tokyo, Japan, 1986

[6] R. B. Ives, “A minimal rule A.I. expert system for real-time classificationof natural spoken languages”, In Proceedings 2nd Annual Artificial Intelligence and Advanced Computer Technology Conference, Long Beach, CA, April-May 1986

[7] M. Sugiyama, “Automatic language recognition using acoustic features”,In Proceedings 1991 IEEE International Conference on Acoustics, Speech and Signal Processing, Toronto, Canada, May 1991