iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 3 **Issue:** II **Month of publication:** February 2015

DOI:

Volume 3 Issue II, February 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

An Introduction to Lid Systems

Niraj Kumar Singh[1], Prof. Anoop Singh Poonia[2]
[1]Research Scholar (M.Tech*), [2]Professor, E & C Engineering
Vivekananda Global University, Jaipur

I. INTRODUCTION

The main task of automatic Language Identification (LID) is to quickly and accurately identify the language being spoken. Language identification has numerous applications in a wide range of multi-lingual services e.g. LID systems used to route an incoming telephone call to a human switchboard operator fluent in the corresponding language. Also, the LID application can be used to handle emergency calls. An LID system can serve as a front end for multi-lingual translation software. Humans are currently the most accurate language identification systems in the world. With a short period of training, people are able to identify a language within seconds of hearing an utterance. Even if one is not able to speak in a particular language then also he or she can relate the unknown language to a particular known language by observing the similarities in utterance. So the question inevitably arises as to what the benefits are gained by making Language Identification an automatic process that can be performed by a machine. The obvious answers are the reduced costs and less training time associated with automated LID systems. For multi-language identification tasks, people involved have to be fluent in all languages under consideration or else they have to be trained properly that they can properly distinguish between the languages. This consumes a lot of time and money. On the other hand, an LID system can be trained once and then run on multiple machines simultaneously in order to correctly identify a particular language from a set of languages.

II. DEVELOPMENT OF LID SYSTEMS

The development and use of automatic speaker recognition systems can be traced back to the late 1980's. In 1995, D.A.Reynolds et.al [3-5] proposed the use of GMM speaker models and adapted GMM models i.e. UBM in speaker identification and verification techniques. The pioneering work in automatic LID systems was carried out by M.A.Zissman in 1996 [15]. He presented a comparative study between the different techniques used in automatic LID systems. He compared the different approaches for LID systems such as GMM, PRLM (Phone Recognition Language Modelling) and PPRLM (Parallel PRLM) [15].

In 2000, E.Wong et.al [14] proposed an approach which used GMM-UBM for language modelling as a speed enhanced alternative to the standard GMM system. J.Navratil [16] proposed a phonotactic-acoustic LID system which used a single-multilingual HMM based phonetic recognizer in 2001. E.Wong et.al [17] utilised Vocal Tract Length Normalization for robust Language Identification in 2002. In 2005, H.Li et.al [18] proposed a phonotactic language model for spoken language identification. R.Tong et.al [19] integrated acoustic, prosodic and phonotactic features for spoken language identification in 2006.

A.Ziaei et.al [20] proposed a technique to enhance spectral features which are used in spoken language identification in 2008. Three novel features based on spectrum were used, in combination with MFCC and prosody features to improve language identification accuracy. These features are spectral centroid, Renyi entropy and Shannon entropy. In 2008 only, B.Yin et.al [21] presented improvements on Hierarchical Language Identification based on Automatic Language clustering. Crossing Likelihood Ratio and Kullback-Leibler distance measures were introduced for faster and more accurate clustering. A novel feature selection scheme based on fusion was proposed in this paper to incorporate multiple features at each classification level [21]. K.C.Sim et.al [22] presented a paper on Acoustic Diversification of the Front End of the LID system. They described a new approach for building a PPRLM system that aims at improving the acoustic diversification among its parallel subsystems by using multiple acoustic models. In 2009, A.Ziaei et.al [7] proposed a new approach for spoken language identification based on Sequence Kernel SVM's. The proposed system consists of a mapping matrix and a
back-end classifier of SVMs as its main parts, located in series after the GMM-LM system. While the mapping matrix maps the language model's output vectors to a new space in which the languages are more separable than before, each SVM in the SVM bank-end classifier separates one language from the others. A new sequence kernel is used for each SVM in the bank-end classifier [7]. In 2010, S.A.Al-Dubaee et.al [23] proposed a language identification technique using Wavelet Transform and Artificial Neural Network (ANN). The classifier used was a threelayered feed-forward artificial neural network and the feature vector was formed by calculating the wavelet coefficients [23]. Y.Xu et.al [6] proposed some methods to improve GMM for language identification in a

paper published in 2010. The score vectors were calculated with Linear Discriminative Analysis (LDA) technique. It was found that with LDA processing approach, the LID system achieved an average accuracy rate of 80% [6]. A.Dustor et.al [10] used GMM modelling to identify spoken languages of 15 different language models. They calculated the results using normalized score techniques and presented the same in a paper published in 2010 [10].

Recently in 2011, F.S.Richardson et.al [24] proposed a new technique for high language identification. They used Nuisance Attribute Projection (NAP) for high level language identification. Also A.Sangwan et.al [25] identified unknown languages using a combined articulatory prosody framework. They proposed a new verification system which was based on dynamic Phonological Features (PF) representation. They achieved an identification rate of 86.6%. This is indeed an achievement when compared to the 56.6% identification rate of the GMM-UBM system proposed by E.Wong et.al [14] in 2000. Clearly, there have been tremendous improvements in the field of Automatic Language Identification over the past decade.

Over the years, many researchers in the field of speech processing and language identification have suggested new techniques in feature extraction and verification which presented a significant development over the traditional approaches. Many new back-end classifiers were also suggested such as Hybrid Hidden Markov Model Neural Network (HHMM-NN), Semi- Continuous Hidden Markov Models (SC-HMM), Dynamic Hidden Markov Network (DHMnet) [9], Multi-Layer KOHONEN Self-Organizing Feature Maps (MLKSFM) etc. A detailed note of the different approaches used in Automatic Language Identification over the past few years is reflected in this paper.

III. SPEECH INFORMATION FOR LID

There is a variety of information that humans and machines can use to distinguish one language from another. The languages of the world differ from one another along many dimensions which have been classified as linguistic categories. At low levels of classification, speech features such as acoustic, phonetic, phonotactic and prosodic information are widely used in LID tasks. When higher levels of classification are required, then differences between languages are determined using morphology and sentence syntax. The various types of information contained in speech utterances are described below:-

A. Acoustic Information: Human speech is basically longitudinal pressure waves. Different speech events can be distinguished according to amplitude and frequency components of the speech waves. These make up the acoustic information class derived from a speech utterance for a particular language.

B. Phonotactic Information: They are a finite set of meaningful sounds that can be produced physically by humans in a particular language. They are also referred to as phonemes.

C. Prosodic Information: Prosody is one of the key components in human auditory perception. It consists of pitch (fundamental frequency of utterance) which relates to the tone, intensity which relates to stress and duration sequence which relates to the rhythm of utterance.

D. Morphological Information: Morphology is the field of linguistics that studies the internal structure of words. A word is divided into two parts: lexicon and a word root. Morphological information is used for higher levels of classification such as determining the emotion or gender of speaker.

E. Syntactical Information: Syntax refers to the order or correct sequence of words needed to form a sentence.

In this project, we are using the acoustic and phonetic information contained in speech. Through this approach we aim at capturing the essential differences between languages by modelling the distributions of spectral features directly.

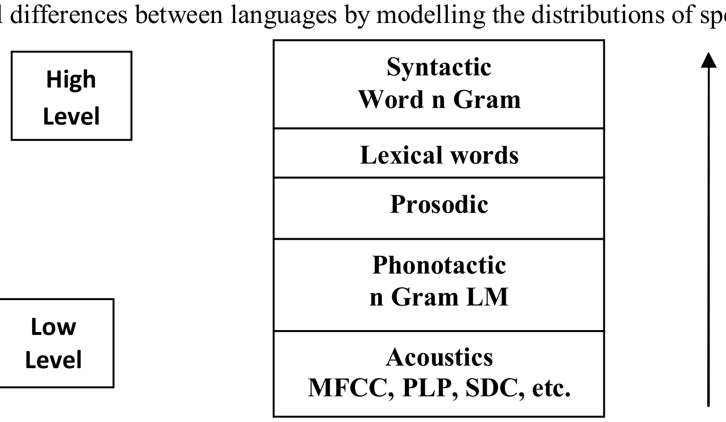


Fig. 1 : The different levels of LID Features [1]

IV. OVERVIEW OF LID SYSTEMS

A standard LID system is divided into two sections: the front-end and the back-end. The front- end extracts a sequence of feature vectors thus parameterizing the speech waveform uttered by a speaker in a particular language. The main purpose of the parameterization process is to extract the most relevant information from the speech waveform and discard as much of the redundant information as possible. In the case of language identification, the parameterization technique removes the speaker and noise dependent properties from the input speech and emphasizes upon the characteristics of the speech waveforms that are most useful for discriminating between different languages. We use the parameterization technique of Mel Frequency Cepstral Coefficients (MFCCs) to transform the speech sequence into a set of feature vectors. The back- end contains the set of language models λL where L is the number of languages. The back-end performs the model training and language identification tasks. In the trainingphase, the feature vectors from the front-end are used to train a separate model λL for each language L to be recognized by the system. We use GMM-UBM (Gaussian Mixture Model-Universal Background Model) as the training model.

V. CONCLUSION

Prosody is the part of speech where rhythm, stress, and intonation are reflected. In language identification tasks, these characteristics are assumed to be language dependent, and thus the language can be identified from them. Voice based biometric systems may prove to be the only feasible approach for remote access control. This novel approach is based on continuous approximations of the prosodic contours contained in a pseudo-syllabic segment of speech. The field of automatic language identification is relatively new and it is progressing at a fast pace. Many new feature extraction and classification techniques have been developed which will increase the identification rate significantly. The main features of LID system are as aforesaid below:

A. The system should not be biased towards any particular language.
B. The computation time should be short i.e. the system cannot be too complex.
C. Increasing the number of target languages or decreasing the time duration of test speech utterance should not degrade system performance.
D. The LID system developed should be robust to speaker and channel variations. The LID system used in this project has been developed keeping the above important points under due consideration.

REFERENCES

[1] E. Ambikairajah, H. Li, L.Wang, B. Yin and V. Sethu, "Language identification: a tutorial," *IEEE Circuits and Systems Magazine*, Second Quarter, pp. 82 – 108,May 2011.

[2] D. Reynolds, "Gaussian mixture models," MIT Lincoln Laboratory, Massachusetts,2002.

[3] D. A. Reynolds and R. C. Rose, "Robust text independent speaker identification using Gaussian mixture models," *IEEE Trans. Speech Audio Processing*, vol.3, no.1, pp. 72 – 83, Jan. 1995.

[4] D. A. Reynolds, "Speaker identification and verification using Gaussian mixture speaker models," *Speech Commun.* , vol. 17, pp. 91 – 108, Mar. 1995.

[5] D. A. Reynolds, T. F. Quatieri and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Processing*, vol. 10, no. 3, pp. 19 – 41,2000.

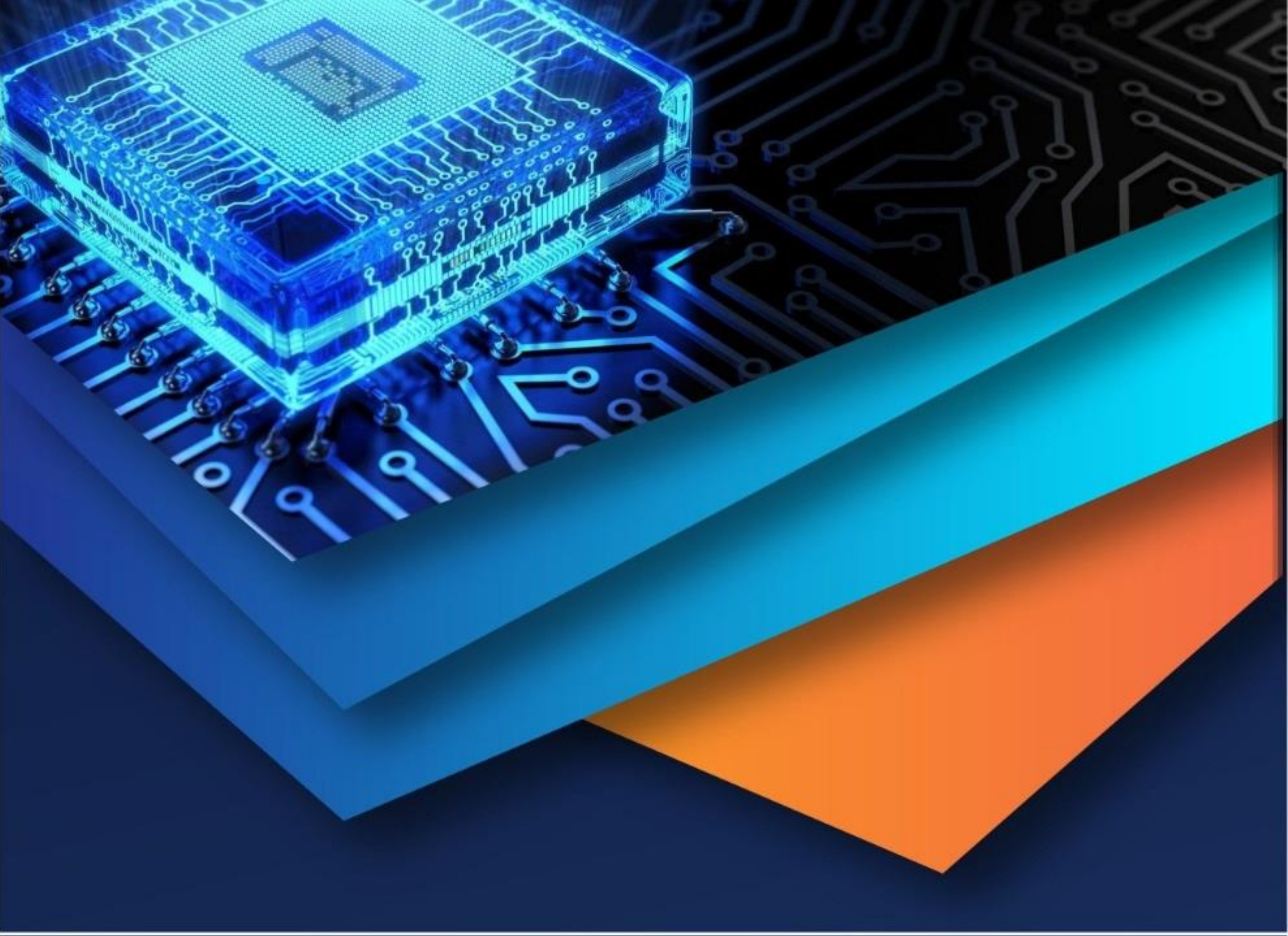
[6] Y. Xu, J. Yang and J. Chen, "Methods to improve Gaussian mixture model for language identification," in *Proc. Int. Conf. Measuring Technology and Mechatronics Automation*, 2010, pp. 656 – 659.

[7] A. Ziaei, S. M. Ahadi, H. Yeganeh and S. M. Mirrezai, "A new approach for spoken language identification based on sequence kernel SVMs," *IEEE Digital Signal Processing Journal*, 2009.

[8] A. Ziaei, S. M. Ahadi, H. Yeganeh and S. M. Mirrezaie, "Spoken languageidentification using a new sequence kernel-based SVM back-end classifier," *IEEE Digital Signal Processing Journal*, pp. 324 – 329, 2008.

[9] K. Markov and S. Nakamura, "Language identification with dynamic hidden Markov network," in *Proc. ICASSP*, 2008, pp. 4233 – 4236.

[10] A. Dustor and P. Szwarc, "Spoken language identification based on GMM models," in *Proc. Int. Conf. Signals and Electronic Systems (ICSES)*, Sept. 2010, pp. 105 – 108.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429