NATIVE LANGUAGE IDENTIFICATION BASED ON ENGLISH ACCENT

G. Radha Krishna
Electronics & Communication Engineering
VNRVJIET
Hyderabad, Telengana, India
guntur_radhakrishna@yahoo.co.in

R. Krishnan
Adjunct Faculty
Amritha University
Coimbatore, India
drrkdrrk@gmail.com

Abstract

Present work is aimed at investigating the influence of mother tongue (L1) of a South Indian speaker on a second language (L2). Second language can be a dominant local language, national language in India i.e., Hindi or a connecting language English. In the current study, L2 is a short discourse in English. Cepstral and prosodic features were used as in Language Identification (LID) to distinguish languages. Both perceptual features and acoustic prosodic features were employed to train Gaussian Mixture Models (GMMs). Studies are carried out with each of the South Indian languages Telugu, Tamil and Kannada as L1. Results showed accuracies upto 85%. Difference in prosodic features of non-native speech is found to be a useful tool for identifying the native state of a polyglot.

1 Introduction

A method of finding the mother tongue adds flexibility to a Text Independent Automatic Speaker Recognition (ASR) system [1] [2]. A possible implementation of this task can be an estimation of the influence of speaker's native language (L1) on a foreign Language (L2). In general, multilingual speakers do not acquire a second language (L2) thoroughly and speech by a particular group of non-native speakers has a distinct 'foreign accent', since they resort to similar type of pronunciation errors. Speaker nativeness or ethnicity can be identified by studying the acoustic and prosodic aspects that remain native like or become most prominent during a discourse [3]. It is observed that non-native speakers inadvertently carry phonemic details from L1 to L2. Studies indicate that Phonetic correlates of accent in Indian English are found in Indian languages [4]. The application areas of mother tongue identification ranges from Intelligence to adaptation in ASR and Automatic Speaker Verification System (ASV), which may require compensation for accent mismatch [5]. A user friendly ASV system for establishing speaker nativeness by establishing the Mother Tongue Influence (MTI) is attempted in this work.

For text-independent nativity recognition, it is possible to create models, which captures the sequential statistics of more basic units in each of the languages. For example, the phonemes or broad categories of phonemes. Modeling approaches can be on the lines of two well-known tasks: Language Identification (LID) and Automatic Speaker Verification/Identification [6]. Some of the successful approaches in this direction include LID using MFCC for Text Independent speaker recognition in multilingual environment and Regional and Ethnic group recognition using telephone speech in Birmingham.

Indian languages are among the less researched languages. ASR Systems are not yet launched into the Indian market at full level. In most of the Indian states, at least two languages are spoken apart from the local official language. This includes English, and a language of the neighbouring province. Popular languages from three South Indian states which are Telugu (ISO 639-3 tel), Tamil (ISO 639-3 tam), and Kannada (ISO 639-3 kan) are chosen for this study. Previous work on

D S Sharma, R Sangal and J D Pawar. Proc. of the 11th Intl. Conference on Natural Language Processing, pages 63–67,
Goa, India. December 2014. ©2014 NLP Association of India (NLPAI)

Nativity identification involved in using both native and non-native acoustic phone models where mapping of phone set from non-native to native language were investigated [4]. In present work, detection of L1 has been attempted by estimating Mother Tongue Influence (MTI) on L2. Language models based on GMM technique were built for each language with a total duration of around 60 minutes per language. The procedure detailed in [7] is followed for this purpose. These models represent the vocal tract at the instance of articulation and will be able to distinguish phonetic features. This can help to identify the speaker's mother tongue which in turn gives the origin of the speaker. A series of experiments are conducted to prove the above approach. Test utterances used were English utterances from Speakers, belonging to the three South Indian regions with above languages as mother tongue. The results for establishing the nativity are promising.

The organization of the paper is as follows: In Section 2, Corpus collection is described. The Modelling technics employed in our experiments are given in Section 3. Results and discussion are contained in Section 4. Finally, Conclusion and scope for future work is given in Section 5.

2 Corpus Description

The speech corpus is collected based on the availability of native speakers of the particular language. Building up of the home grown corpus is described below. The speakers are separated into two groups: training and testing set. Speech samples are collected from native Speakers belonging to the states of Andhra Pradesh, Tamil Nadu or Karnataka whose mother tongue are respectively Telugu [TEL], Tamil [TAM] or Kannada [KAN]. This constituted the training set. The speakers are so chosen that they are not from places bordering other states. This ensures that dialectal variation is avoided in the training set. A total of 3600 seconds of speech corpus is developed for each of the three languages. The details are given in Tables 1 and 2.

Recording is carried out with text material from general topics related to Personality development and with the speakers under unstressed conditions. A different subsets of speakers who are capable of speaking English in addition to the above mentioned mother tongues are chosen as the testing set. Thus the testing database consisted of English utterance of the speakers with one of the three languages Telugu, Tamil or Kannada as mother tongue. It is ensured that Gender weightages are almost equally distributed in both the training and testing sets. The test utterances, which are English samples are recorded under similar conditions as training speech samples. The details of speakers of test set are detailed in Section 4. Each of the test sample is recorded for a duration of 90 Seconds. These details are shown in following Table 3

Table 1: Distribution of Training Set

Language		TEL	TAM	KAN
No. of speakers	M	5	3	4
	F	4	3	4
No. of minutes	M	30	35	25
	F	30	25	35

Table 2: Speaker Proficiency in other languages

Language	Male	Female
TEL	HINDI	NIL
TAM	NIL	ENGLISH
KAN	HINDI,ENGLIH	HINDI,ENGLISH

Table 3: Distribution of Testing Set

Language		TEL	TAM	KAN
No. of speakers	M	7	7	4
	F	7	5	8
No. of Seconds	M	30-90	30-90	30-90
	F	30-90	30-90	30-90

3. Experiments

3.1 System building: According to [6], Language identification is related to speaker-independent speech recognition and speaker identification. It is practically easy to train phoneme models than training models of entire language. Though they are found to outperform those based on stochastic models, the phonemic approach has the following drawback. It needs phonemically labeled data in each of the target languages for use during the training. The difference among languages, apart from their prosody lies in their short-term acoustic characteristics. Indian languages share many phones among themselves. Since there are many variants of the same phoneme, we need to consider the acoustic similarities of these phones. Combination of phonetic and acoustic similarities can decide a particular mother tongue [3]. For text-independent language recognition, it is generally not feasible to construct word models in each of the target languages [8]. So, models based on the sequential statistics of fundamental units in each of the languages are employed. Text independent recognizers use Gaussian mixture models (GMMs) to model the language dependent information. The modeling technic deciding the acoustic vectors should be multimodal, to represent the pronunciation variations of the similar phonemes in various languages. The language model used in this particular study is a GMM model of Mel Frequency Cepstral Coefficients MFCCs [9]. Following block diagram (Fig. 1) illustrates the implementation of above steps in the frame work of a Speaker Recognition system. The system is an acoustic information based LID system for which the proposed Foreign Accent Identification system is a special case.

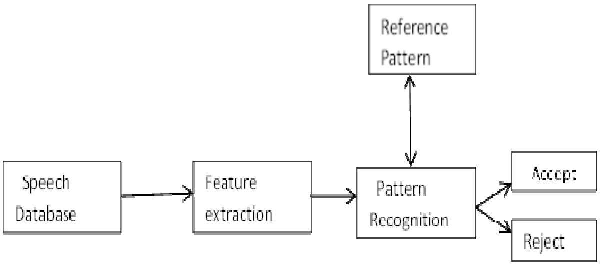

Figure 1: Speaker Recognition system for nativity identification

3.2 Spectral features for Language Identification:

Present day Speaker recognition systems rely on low-level acoustic information [10]. Studies indicate that a cohesive representation of the acoustic signal is possible by using a set of mel-frequency cepstral coefficients (MFCCs) which emulates the functioning of human perception. MFCCs are cepstral domain representation of the production system. MFCCs are 13 dimensional vectors which help in several speech engineering applications. The speech signal is converted into a set of perceptual coefficients represented by a 13 dimensional MFCC feature vector. After collecting the multilingual speech set, acoustic model parameters are estimated from the training data in each language. The extraction and selection of the parametric representation of acoustic signals is critical in developing any speaker recognition system. Cepstral features capture the underlying acoustic characteristics of the signal. They characterize not only the vocal tract of a Speaker but also the prevailing characteristics of the vocal tract system of a phoneme. In conclusion, MFCCs provide information about the phonetic content of the language. Hence, we used MFCC coefficients as feature vectors to model the phonetic information.

3.3 Experiments based on stochastic models:

GMMs are famous classification technique which helps to cluster the input data into a pre-determined specifications about clusters. GMMs are a supervised technique which is efficient in classifying multi-dimensional data. The main purpose of using the Gaussian mixture models (GMM) in pattern recognition stage is because of its computational efficiency. Moreover, the model is well understood, and is most suitable for text-independent applications. It is robust against the temporal variations of the speech, and can model distribution of acoustic variations from a speech sample [7][9]. The GMM technique lies midway between a parametric and non-parametric density model. Similar to a parametric model it has structure and parameters that control the behavior of the density in known ways. It also has no constraints about the type of data distribution [7]. The GMM has the freedom to allow arbitrary density modeling, like a non-parametric model. In the present investigation, the Gaussian components can be considered to be modeling the broad phonetic sounds that characterize a person’s voice. The proposed Mother Tongue Identification system is based on the statistical modeling of Gaussian mixtures [11].

4. Results and Discussion

In the testing phase, speech samples from a set of speakers with wide ranging geographical distribution within a state are collected. The speakers in test set are all educated, with at least graduation. Teachers of English language, convent educated are avoided in the test set. Most of the speakers have the ability to speak one or more local languages apart from English, representing a truly multilingual scenario. These speakers are fluent in English as well as in their mother tongue. The test samples are modeled similarly as training samples and compared with three baseline Language models developed in the earlier training phase. Distance measures are computed between the GMM mean of each language model and that of the test utterances of MFCCs parameters derived from the test utterance. Confusion matrix of pair-wise mother tongue identification task is performed and the results are presented in Table 4.

Table 4. Confusion matrix of pair-wise MTI task.
(a) Between Telugu and Tamil

(i) Cepstral features

30 Sec	TEL	KAN
TEL	80%	20%
TAM	20%	80%

(ii) Acoustic-prosodic features

60Sec	TEL	KAN
TEL	85%	15%
TAM	20%	80%

(b) Between Telugu and Kannada

(i) Cepstral features

30 Sec	TEL	KAN
TEL	80%	20%
KAN	20%	80%

(ii) Acoustic-prosodic features

6 0 Sec	TEL	KAN
TEL	80%	20%
KAN	20%	80%

(c) Between Tamil and Kannada

(i) Cepstral features

30 Sec	TEL	KAN
TEL	80%	20%
TAM	20%	80%

(ii) Acoustic-prosodic features

6 0 Sec	TEL	KAN
TEL	80%	20%
KAN	20%	80%

5. Conclusions and Future scope

An Automatic Speaker Recognition system for identification of mother tongue and thus the native state of the speaker is implemented successfully. Confusion is observed between Kannada and Tamil speakers. This confusion is found to be less when Acoustic prosodic features are introduced. We have proposed an effective approach to identify MTI in multilingual scenario by following the techniques available in Language and Speaker Identification. A general purpose solution is proposed with a multilingual acoustic model. Further improvements can be made by including prosodic features and also covering techniques such as inclusion of SDC features and also the i-vector paradigm. Most important advances in future systems will be in the study of acoustic-phonetics, speech perception, linguistics, and psychoacoustics [7]. Next generation systems need to have a way of representing, storing, and retrieving various knowledge resources required for natural conversation particularly for countries like India. With the same training and testing procedures, apart from English and other regional languages, national language Hindi can be modeled and influence of any particular language on it can also be studied.

Acknowledgments

The authors would like to acknowledge the cooperation of the staff and students of VNRVJIET, who co-operated readily by providing their voice samples. We profusely thank all these speakers for their kind co-operation towards carrying out this research. Special thanks are due to the Scientists of Speech and Vision Laboratory of IIIT, Hyderabad for their timely and invaluable advice.

References

[1] G. Doddington, P. Dalsgaard, B. Lindberg, H. Benner, and Z. Tan, “Speaker recognition based on idiolectal differences between speakers”, in Proc. EUROSPEECH, pp. 2521–2524, Aalborg, Denmark, Sep. 2001.

[2] A. Maier et.al., “Combined Acoustic and Pronunciation Modeling for Non-Native Speech Recognition” Interspeech 2007, pp1449-1452.

[3] R.Todd, “On Non-Native Speaker Prosody: Identifying ‘Just-Noticeable-Differences’ of Speaker-Ethnicity”, Proceedings of the 1st International Conference on Speech Prosody, 2002

[4] E. Shriberg, L. Ferrer, S. Kajarekar, N. Scheffer, A. Stolcke, and M. Akbacak, “Detecting non-native speech using speaker recognition approaches”, in Proceedings IEEE Odyssey-08 Speaker and Language Recognition Workshop, Stellenbosch, South Africa, Jan. 2008.

[5] Sethserey et.al. Speech Modulation Features for Robust Nonnative Speech Accent Detection, Interspeech-2011

[6] “Multi Level Implicit features for Language and Speaker Recognition”, Ph.D. Thesis, Leena Mary, Department of Computer Science, Indian Institute of Technology Madras, India ,June 2006.

[7] D. A. Reynolds, ” Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models”, IEEE Transactions on Speech and Audio Processing Vol.3,No.1,1995,72-83.

[8] A. Maier et.al “A Language Independent Feature Set for the Automatic Evaluation of Prosody ” Interspeech 2009.

[9] N. Scheffer, L. Ferrer, Martin Graciarena, S. Kajarekar, E. S. Stolcke, “ The SRI NIST 2010 Speaker Recognition Evaluation System ”.

[10] D.A. Reynolds, T.F. Quatieri and R.B.Dunn, “Speaker Verification using adapted Gaussian mixture models”, Digital Signal Processing vol 10, pp19-41, 2000.

[11] J. Cheng, N. Bojja, X. Chen “ Automatic Accent Quantification of Indian Speakers of English” Interspeech 2011, pp2574-2578.