Non-native Accent Partitioning for
Speakers of Indian Regional Languages

G Radha Krishna
Dept. of E.C.E,VNRVJIET
Hyderabad, India
gunturrkr@gmail.com

R Krishnan
Amritha University
Coimbattore, India
drrkdrrk@gmail.com

V K Mittal
K L University
Vijayawada, India
drvinaykrmittal@gmail.com

Abstract

Acoustic features extracted from the speech signal can help in identifying speaker related multiple information such as *geographical origin*, *regional accent* and *nativity*. In this paper, classification of native speakers of South Indian languages is carried out based upon the accent of their non-native language, i.e., English. Four South Indian languages: *Kannada*, *Malayalam*, *Tamil*, and *Telugu* are examined. A database of English speech from the native speakers of these languages, along with the native language speech data was collected, from a non-overlapping set of speakers. Segment level acoustic features *Mel-frequency cepstral coefficients (MFCCs)* and F_0 are used. Accent partitioning of non-native English speech data is carried out using multiple classifiers: *k-nearest neighbour (KNN)*, *linear discriminant analysis (LDA)* and *support vector machine (SVM)*, for validation and comparison of results. Classification accuracies of 86.6% are observed using KNN, and 89.2% or more than 90% using SVM classifier. A study of acoustic feature F_0 contour, related to L_2 intonation, showed that native speakers of Kannada language are quite distinct as compared to those of *Tamil* or *Telugu* languages. It is also observed that identification of *Malayalam* and *Kannada* speakers from their English speech accent is relatively easier than *Telugu* or *Tamil* speakers.

1 Introduction

Identification of speakers, classification of their dialectal zones is important in a multilingual country like India (Bhattacharjee and Sarmah, 2012). Speaker uniqueness is manifested in both anatomical and learned traits. When the context is constrained, speaker characteristics can be used reliably to identify individuals (Arslan and Hansen, 1996). The *accent* is one of the glaring indications of linguistic and social background of a speaker. Studying the characteristics of dialect on a phonetic or phonemic level belongs to accent recognition (Mittal et al., 2014). Earlier studies have concluded that native language (L_1) affects the speaker's traits of their second language (L_2) (Ghorbani et al., 2018; Graham and Post, 2018). Analysis and classification of utterances that belong to specific groups of learners is the main objective of *Native Language Identification* (NLI) (Nisioi, 2015). However, there is very little research on the question of accuracy with which accent features can be used to identify a speaker's *regional* or *ethnic origin* (Harper and Maxwell, 2008). A solution to the problem of regional accent classification across English speaking South Indians is attempted in the present research, using a specifically developed corpus.

Discriminative classifiers based on characterizing acoustic differences across foreign accents can be employed to direct an accent dependent recognition system (Omar and Pelecanos, 2010; Ikeno and Hansen, 2006). Systems with an automatic evaluation of non-native speech, which includes characteristics of the mother tongue will have better performance over similar algorithms that depend upon target languages (Qian et al., 2017). This is particularly true when the text uttered is unknown. Native listeners are mostly aware of the speaker's regional accent and also the social or geographical subgroup within the region (Hanani et al., 2013). Automatic speaker characterization is vital in real-world applications and the advantages are widely open (Zampieri et al., 2017; Krishna and Krishnan, 2014).

Pattern recognition approach of collecting data, extracting suitable features, and training classification module using machine learning is a powerful tool in applications like Computer-Assisted-Pronunciation-Training (CAPT) programs. Acoustic descriptors are critical in tasks

Table 1: Summary of data used for training and testing: (a) attributes (b) values for *training set* and (c) values for *testing set*

(a) *Attributes*	(b) *Training set*	(c) *Test set*
Total number of speakers	60	75
Speakers per language group (KAN, MAL*, TAM, TEL)	20	25
Speech Duration per speaker	300 sec	60 sec

Note: *MAL-Malayalam data set is used only in tests related to cepstral features.

such as sound Classification (Day and Nandi, 2007). State-of-the-art Accent Identification (AID) systems widely rely on spectral acoustic distribution for modeling the pronunciation. In applications like accent recognition, features distinguishing different phonemes of a language will be useful (Neumeyer et al., 2000). Language-specific differences in phonological development might be related to differences in phoneme and phoneme sequence frequency across languages (Ikeno and Hansen, 2006). Such variations are also represented by the intonation patterns of individuals (Mary and Yegnanarayana, 2008; Li et al., 2017). Apart from cepstral features that capture underlying acoustic characteristics, information from higher-level prosodic traits (Doddington, 2001; MALMASI and DRAS, 2017) were examined in the present study.

English is the most widely spoken second language in India and elsewhere in the world (Saha and Mandal, 2015; Guntur et al., 2018). Indian English has several varieties with their specific accents and phonological features and often a distinct lexicon. Research on spoken English of Indian speakers is urgently needed from a multi-disciplinary perspective (Cheng et al., 2013; Krishna et al., 2019). Present work is aimed at comparing the acoustic properties that are likely to differ between English accents different groups of South Indian language of speakers. The non-native prosodic traits are a hindrance to proficiency in a second language (L_2), and also to the mutual understanding. Present work also examines the local prosodic changes in the non-native English speech, without incorporating any phonol-

ogy of the specific languages. The ability to compensate against prosodic deviation during English production can be improved by identifying the articulatory gestures that emphasize the non-native speaker accent.

Table 2: Template of *file naming* for data recording

Native language	Name	Age / Sex	File Name

The paper is organized as follows: Section 2 presents the details of the database, including the recording methodology. Section 3 describes acoustic and prosodic features used in foreign accent recognition. Section 4 describes the classification procedures employed in the NLI experiments. Section 5 gives the details of the experiments and results. Analysis of results of regional accent classification is given in section 6. Section 7 describes the key outcome and contributions. Conclusions drawn are given in Section 8.

2 Data Sets of 4 Indian Regional Languages

The main focus of current research work is on differentiating the regional non-native English accents of speakers, and also describing foreign accent in terms of a common set of fundamental speech attributes. A database has been specifically developed (G.Radha Krishna and Mittal, 2018) with native and non-native speech samples containing utterance by the speakers belonging to language groups Kannada (KAN), Malayalam (MAL), Tamil (TAM), and Telugu (TEL). Table 2 shows the template of file naming process.

2.1 Selection of Regional Languages

Among more than six thousand languages in the world, less than 10% of the languages are spoken by more than 90% of the people. Speakers and learners of the English language constitute a large proportion in countries like India, South Africa, and much of the developing world. India has distinct linguistic communities, each of which shares a common language and culture. English, Hindi and dominant local languages are spoken non-natively by a large number of Indians. In South Indian cities, many people speak at least two second languages. It would be beneficial if speech based systems can store models of all known languages and carry out the task of NLI automatically.

Table 3: Summary of *speaker traits* and related *speech features* (Day and Nandi, 2007).

Speech characteristic	Speaker trait	Speech feature
Lexical, Syntactic (Idiolect, Semantics, Pronunciations, dictions, Idiosyncrasies)	Socio economic Educational status (Language use and sentence construction)	Vocabulary, Word arrangement & grammatical cues.
Prosodic (Rhythm, Intonation, Articulation rate etc.)	Personality type, Parental influences	Durational features. Pitch dynamics, Energy (likely to be Text / time dependent).
Low level acoustic features	Anatomical structure of speaker's vocal apparatus	Short-time spectrum, Predictor coefficients, Intensity, Pitch.

Table 4: Major *text-independent features* used in *prosodic analysis*.

Prosodic features	Factors that influences speech
Dynamics of F_0 contour	Identity of sound unit, its position from phrase, word; Speaking style; Intonation rules; Type of sentence (Interrogative, Declarative)
Intonation, Rhythm, Stress	Attitudinal, Accentual, Discourse, Grammatical

2.2 Speech Corpus Recording Methodology

The details of speech corpus developed for each of the languages is shown in Table 1. Native speech utterances of 20 speakers from each of the native language groups KAN, TAM, and TEL, each with a duration of 300 seconds formed the training set. English test samples for a duration of 60 seconds were collected from 25 speakers belonging to each of the four groups KAN, MAL, TAM, and TEL. As the sufficient number of native speakers of MAL are not readily available, it is included in the testing set only. The test utterances were recorded under identical conditions as training speech samples and there is no overlap between training and testing sets with respect to speakers and sentences. Each of the test samples is recorded for a duration of 60 seconds. The non-native English speech samples are collected from a set of speakers with nearly uniform geographical distribution within a region with an educational background of at least graduation, but who do not use English routinely.

Recordings of speakers were made in quiet office room conditions using Logitech h110 microphone and waveforms are sampled at a rate of 16 kHz. The recordings were made in a laboratory environment with written text, with negligible reverberation. The participants were asked to read aloud passages of a text from general topics. For applications like screening of non-native speech, read data can be used for both training and testing (Schuller et al., 2013). It is ensured that Gender weightages are equally distributed in training as well as testing data sets. The speakers in the training set are considered representative of the regional languages KAN, TAM and TEL. However, for testing set speakers of Malayalam were also included. These speakers are so chosen from language heartlands. The speakers in the test set are considered potential users of future systems augmented with automatic Accent Identification (AID) capability.

3 Features for Non-native Accent Partitioning

Understanding similar variations in foreign accents is a crucial factor for the development of an NLI system. The dominant articulatory traits of different languages are different (Koreman, 2018). In applications like accent recognition, features distinguishing different phonemes of a language will be useful (Li et al., 2013). The acoustic signature or the **voice individuality** of the speech signal are available as differences in transformations occurring at semantic, linguistic articulatory, and acoustic levels. Out of all the factors affecting speech, accent is a week factor in the sense that speech variation is not as evident as that due to speaker/gender.

Language-specific differences in phonological development might be related to differences in phoneme and phoneme sequence frequency across languages (Graham and Post, 2018). Speakers of the second language (SL) are expected to import certain patterns from their native language (NL)
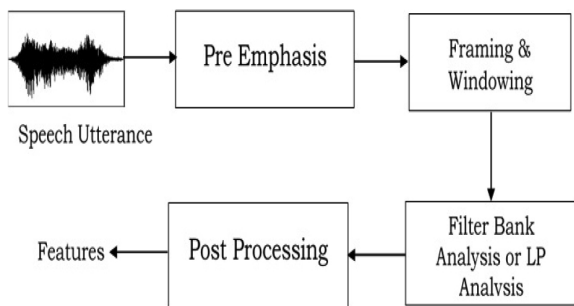
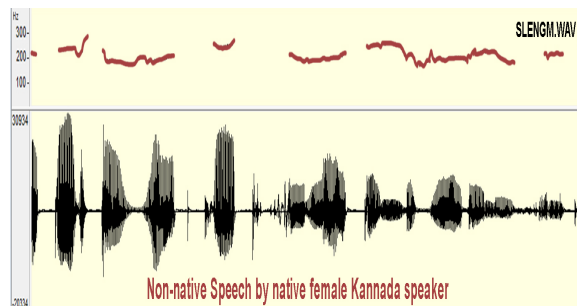Figure 1: Front end *signal processing* for feature extraction

Figure 2: Waveform and Pitch contour of *non-native English speech* by female Kannada speaker

which are audible in SL. The influence of the surrounding speech prosody on new-born cry melody has been shown (Monnin and Loevenbruck, 2010). The non-native speech detection is thus very challenging .

Characterization of a foreign accent is mostly based on either auditory analysis or manual transcriptions of deviations. The auditory spectrum is consistent with several phenomena observed in speech perception and is useful in automatic speaker independent speech recognition. Features used for nonnativeness detection include cepstral vectors, phone strings and a variety of prosodic features, but when used alone, systems based on acoustic features perform better (Shriberg et al., 2005). We can consider acoustic features, which are proxy of phonetic reproduction as acoustic-phonetic features (Li et al., 2013).

3.1 Acoustic Features

Earlier investigations on text-independent non-native speech tied to underlying native language structure are based on (i) Global acoustic distribution of phonemes (which requires no language knowledge) (ii) Different intonations corresponding to uniqueness in the manner in which articulators are manipulated. The shape of the vocal tract is manifested in the envelope of the short-time power spectrum (Reynolds and Rose, 1995). The attributes that contain speaker identifiability for machine as well as for humans are of interest (Zheng et al., 2007; Franco et al., 2000).

In this study, acoustic features used for phonetic modeling of the accent differences consists of the cepstral features: Perceptive Linear Prediction Coefficients (PLPs), Linear Predictive Cepstral Coefficients (LPCCs), and MFCCs (Hermansky, 1990; Luengo et al., 2008; Mittal and Yegnanarayana, 2013). The steps followed are shown in Figure

1. Given all the alternative spectral features based on LPC - cepstrum and FFT cepstrum for speaker recognition, MFCCs, give a highly compact representation of the spectral envelope of a sound (López, 2014). The LPCCs are known to capture extra information from a speech that discriminates different languages. The PLPs which take advantage of psychoacoustic principles are robust against noise. A hierarchy of speech characteristics, related speaker traits, and possible speech features are listed in Table 3.

3.2 Prosodic Features

The prosodic structure is a critical aspect of language contact and gives important information related to the speaking habit of a person (Kinnunen and Li, 2010; Farrús et al., 2010). The goal is to capture prosodic idiosyncrasies of speakers belonging to different native languages. Prosodic cues Stress, Rhythm, and Intonation are each complex entities expressed using (i) Pitch (ii) Energy (iii) Duration. Major text-independent features used in prosodic analysis are given in Table 4.

In this study Prosodic statistics were obtained by performing different measurements of *pitch*, which are derived supra segmentally. The power of accent in voice identification is investigated as explained below. A *Generative model* of pronunciation describes what is acceptable, and *Discriminative model* both acceptable and unacceptable pronunciation, and the pronunciation score is the direct output of the classification module.

Non-native prosodic traits limit proficiency in a second language (L_2). Prosodic phenomena located on word level and above, help listeners to structure the speech signal and to process the linguistic content successfully. Table 4 shows some of the features useful for detecting non-native speech without annotation of prosodic events. The
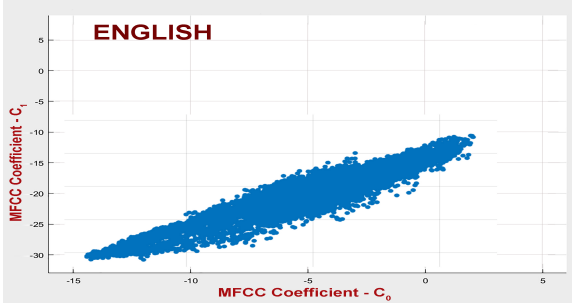
68

Figure 3: Distribution of MFCC Coefficients as a Scatter plot of C_0 versus C_1 for *native ENGLISH* speakers
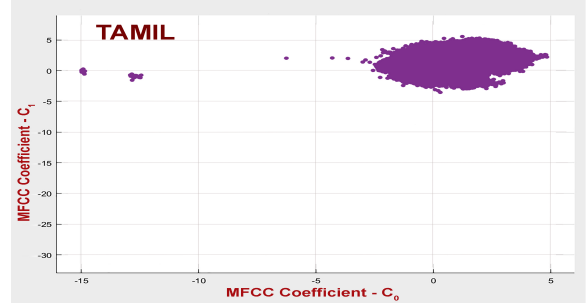

Figure 6: Distribution of MFCC Coefficients C_0 versus C_1 for *English speech by TAMIL speakers*
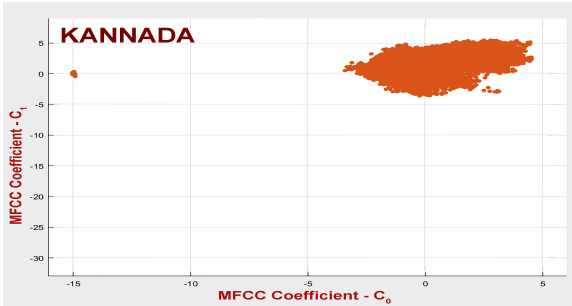

Figure 4: Distribution of MFCC Coefficients C_0 versus C_1 for *English speech by KANNADA speakers*
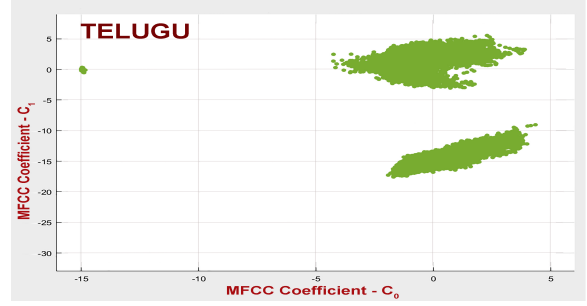

Figure 7: Distribution of MFCC Coefficients C_0 versus C_1 for *English speech by TELUGU speakers*

experiment by Rosenberg to foil a Speaker Verification system says that even an identical twin was unable to imitate the enrolled sibling well enough to get accepted by the system, tells the need to look at learned speaking behaviour.

4 Classification for Non-native Accent Partitioning

Speaker Classification can be conveniently defined as a grouping of speakers speaking in a similar manner, on the basis of acoustic characteristics (Chen et al., 2014). Classification of foreign accents directly from the acoustic features is at-
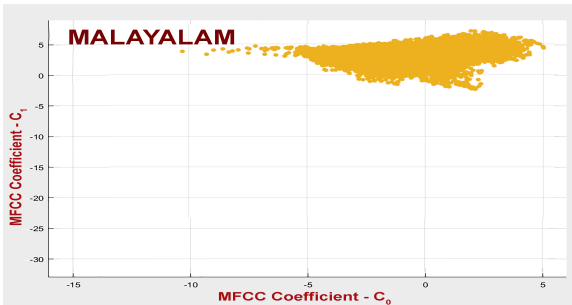

Figure 5: Distribution of MFCC Coefficients C_0 versus C_1 for *English speech by MALAYALAM speakers*

tempted by using a test data set described in Table 1. The role of accent in voice identification is investigated as explained below. There exists a significant overlap between NLI approaches and computational methods for dialect and language identification (LID), and Support Vector Machine (SVM) classifiers are a very good fit for NLI (Zampieri et al., 2017).

4.1 Accent Partitioning using SVM Classifier

SVM is one of the most popular supervised classifiers on a wide range of data sets, which looks for a *maximum-margin hyper plane* for data separation (Wu et al., 2010; Bahari et al., 2013; Campbell et al., 2006). Accuracies of non-native accent classification were studied for the present problem by using the SVM classifier. The speech signal is first processed to extract attributes relevant to the *foreign accent* (Moustroufas and Digalakis, 2007). The most representative acoustic features, the LPCC, the PLP (Li et al., 2013) have been tested but were found to be less efficient. The input to the system is a 13 dimensional MFCC vector consisting of 12 cepstral coefficients and one energy coefficient. Thus the front end for the proposed classification system consisted of only 13 dimensional MFCC vector including C_0.
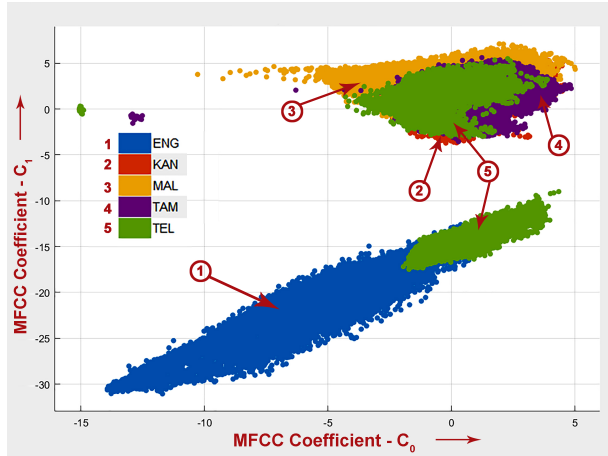
Figure 8: Distribution of MFCC Coefficients C_0 versus C_1 for *non-native English speech by four South Indian language speakers* against *native English speech.*

Figure 9: Confusion Matrix for SVM classification of *South-Indian English including native English.* **Note**: TPR is True Positive Rate, FNR is False Negative Rate.

Table 5: *Non-native Regional English Accent Classification accuracies using (a) k-nearest neighbourhood (KNN), (b) Linear Discriminant (LDA), and (c) SVM*

Classifier	(a) KNN	(b) LDA	(c) SVM
Accuracy	**86.6%**	**82.5%**	**89.2**

4.2 Intonation Analysis

Native traits located at a word and sentence levels help listeners structure the speech signal. In many approaches that apply prosody to either *Language Identification* (LID) or *Speaker Recognition*, extracted features are based on statistics of pitch / energy contour segments or piecewise linear stylization of pitch / energy contours. Intonation is a key expressive factor which can covey the intent of a speaker, contains a lot more information than words and utterance (Ward et al., 2017). *Intonation* is more used than *energy* and *duration* features in the context of prosody. Listeners can discern a speaker's regional accent from intonation alone (Eady and Cooper, 1986; Tepperman and Narayanan, 2008).

Dynamics of F_0 contour corresponding to a sound is influenced by several factors such as the identity of the sound unit, its context, the speaking style of the speaker, intonation rules of the language, type of the sentence, etc. (Arias et al., 2010). The focus was mainly on the pitch since it is one of the most important characteristics of prosody and helps in predicting human intonation rating. These suprasegmental parameters can be used to model non-native English prosody (Hönig et al., 2012). In the present study, the main aim is to ascertain the influence of linguistic background on F_0 across regional varieties of English, future studies are planned to include the aperiodic components of excitation of expressive voices like Noh voice (Mittal and Yegnanarayana, 2015)

5 Experiments and Results

To validate the hypothesis that the accent of the mother tongue is separable, experiments were performed to understand and to calibrate idiolectal differences in the non-native speech samples of the language groups KAN, MAL, TAM and TEL. The corpus is sampled at 16000 samples per second and the bit rate was 32 bits per sample. Silence removal has been implemented using a VAD algorithm (Kinnunen and Li, 2010). The feature vectors are computed over 20 msec windowed frames every 10 msec. Fourier spectra were computed for sequential frames 160 points apart by using a 320 point Hamming window. Finally Cepstral Mean Normalization (CMN) is applied by subtracting the mean value of each feature over the entire utterance. MFCCs are generated by windowing the signal, application of DFT, taking the log of the magnitude and warping the frequencies on Mel scale and finally application of DCT.

5.1 Non-native Accent Classification based upon Acoustic Features

Experiments were performed to establish the differences in the distribution of acoustic features in the non-native speech samples of four language groups KAN, MAL, TAM, and TEL. Graphical il-
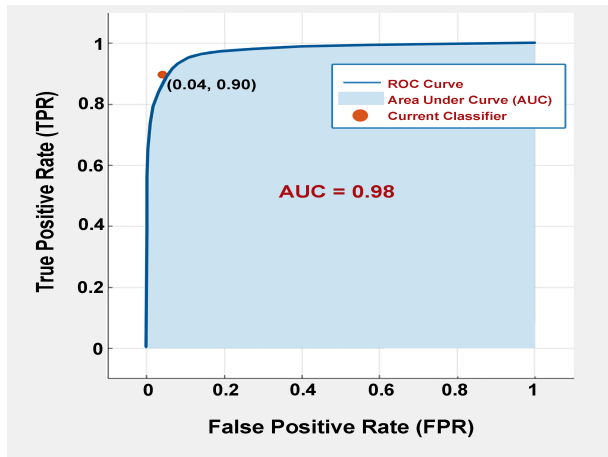
Figure 10: : ROC curve for SVM classification of *Non-native English speech by Kannada speakers.*

Figure 11: Confusion Matrix for SVM classification of *English by speakers of KAN, TAM, and TEL.* **Note**: TPR is True Positive Rate, FNR is False Negative Rate.

lustration of accent partitioning on test data is shown in Figures 3,4,5,6,7, and 8. It indicates that the high classification accuracies are possible in the present task. Classification of foreign accents directly from the acoustic features is attempted, by using data set described in Table 1. Figure 9 shows the confusion matrix for best performing SVM classifier for the five class classification. Figure 11 shows the confusion matrix for the three class classification.

The confusion matrix indicates that the identification rates for Kannada and Tamil language speakers from their non-native English speech can be high compared to that of Telugu native speakers . The Receiver Operating Point Curve (ROC) shown in Figure 10 is a plot of true positive rate as a function of false positive rate, which is very close to the upper left hand corner, indicates that the classifiers can achieve good overall accuracies.

Verification of accent partitioning of non-native speech using a series of classification techniques: k-nearest neighbourhood, and Linear Discriminant Analysis was also implemented. English speech samples of the native speakers of KAN, MAL, TAM, and TEL are tested against standard English speech corpus using TIMIT corpus. The resulting accuracies are 86.6% when a KNN clssifier is used, 82.5% when Discrimination classifier is used, and 89.2% using SVM classifier is used. These results are consolidated in Table 5. Figure 4, and 6 shows the corresponding confusion matrices, obtained during SVM classification.
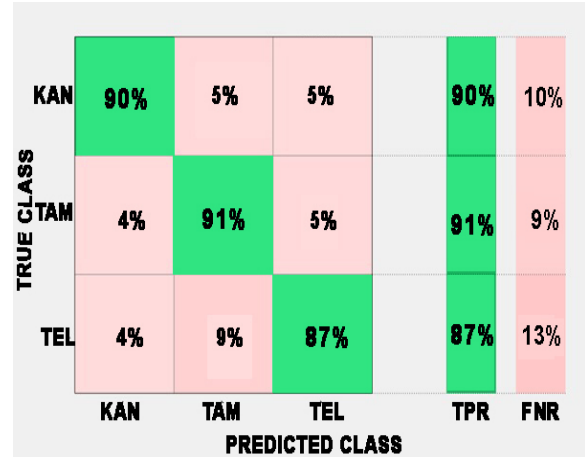
5.2 Foreign Accent Discrimination based upon Prosodic Features

Experiments were conducted on native and non-native speech samples of bilingual and multilingual speakers. The pitch frequency was extracted using the "pitch contour" function of the *Wave Surfer* software, and F_0 data was extracted. Typical waveform showing the non-native speech by a female Kannada speaker and the pitch contour were shown in Figure 2. The speakers in this study were asked to speak in their mother tongue or in English, and 20 exemplars were analysed from each group KAN, TAM, and TEL. In few cases the same speakers have spoken in other Indian language of the neighbouring state.

The difference in F_0 contour between native and non-native speech for speakers from each group has been tested. These results shown in Table 6 clearly indicate that the mean value of non-native pitch is markedly high in the case non-native speakers in all the three groups. The percentage deviation from native language to English speech for a group of 20 speakers in each of the three languages has been estimated and is presented in Table 7. It is evident from the scores presented in Table 7 that the dynamic variation of pitch is the least at 3.7% for the regional variant of KAN speakers, which is significantly less when compared to 9.5%,and 27% corresponding to native TAM and TEL speakers respectively.

Table 6: Mean (μ) and SD (σ) of *Pitch variation* of single speaker from three groups of native speakers when speaking (a) Native Language (NL) (b) English (c) Other South Indian language (OSIL)

L1	LANGUAGE SPOKEN					
	(a) NL		(b) English		(c) OSIL	
	μ	σ	μ	σ	μ	σ
Kan	214	32.2	254	32.3	235	32.4
Tam	227	21.7	248	28.9	230	30.6
Tel	133	21.5	157	22.9	150	26.3

6 Analysis of Results

- Figures 3,4,5, and 8 reveal that the English spoken by native Kannada and Malayalam speakers is distinct than native Tamil or Telugu speakers, when compared to standard English.

- Accent partitioning experiments from a short utterance of 60 seconds of test data, indicates the suitability of the SVM classifier, as can be seen from accuracies shown in Table 5.

- Figure 1 reveals that the English spoken by Telugu native speakers are marginally closer to standard English, compared to that of Kannada and Malayalam language speakers.

- Higher mean values of the non-native pitch shown in Table 6 indicates the accommodation of speakers of all native languages to suit different social groups.

- Table 7 shows that English speakers of Tamil and Telugu would produce statistically significant higher pitch contour deviations than KAN speakers.

7 Key Outcome and Contributions

- A framework to handle the deviations of L_2 influenced by closely related L_1s and to achieve better performance for a given NLI task, even with fewer features is proposed

- Current study is significant when the target languages are linguistically close, and large resources of spoken English are not available

- Prosodic differences across the South Indian English accents has been experimentally illustrated, which is useful in automatic intonation classification for L_2 speech acquisition.

Table 7: *Percentage increase in Standard Deviation* of pitch contour from native language speech to English speech (using two non-overlapping sets of 20 speakers from each native language group Kannada, Tamil, and Telugu).

Language group	Male	Female	Average
Kannada	0.9	6.5	3.7
Tamil	9	10	9.5
Telugu	33	21	27

- Present work helps in accurate recognition of regional accent, that can improve the speech and speaker recognition system performance.

- Distinct pitch pattern variations in non-native English speech by Malayalam, and Kannada speakers compared to that of Tamil and Telugu varieties can help in distinguishing them.

8 Conclusion

It can be concluded that the regional native language classification has been achieved with an accuracy of nearly 90%, by using the acoustic distribution of cepstral features on the four types of non-native South Indian English speech. It is known that systems make more mistakes among regionally close languages. Accent differences among the non-native speakers are reflected as the deviation of L_2 influenced by L_1 on prosodic level. Studies carried out based on intonation distribution indicates that English speaking South Indian groups corresponding to Kannada, Malayalam, Tamil, and Telugu are clearly divided as per their native languages. Prosodic differences in the native and English speech by South Indian speakers were detected without annotation. Present method can potentially be applied to other languages like Hindi, and in addressing the important question of finding a universal feature set for identifying the non-native speech.

Present research is useful in applications such as voice based wireless services like mobile health care, agriculture. Automatic accent characterization can also be applied to fields such as sociolinguistics and speech pathology. Future work can employ different speech styles, and characteristics of speaker population to be carefully scrutinized, and also by including multi-disciplinary information. Further, the results can be extended to separating language families and also for rating L_2 proficiency.

References

Juan Pablo Arias, Nestor Becerra Yoma, and Hiram Vivanco. 2010. Automatic intonation assessment for computer aided language learning. *Speech Communication*, 52(3):254–267.

Levent M. Arslan and John H.L. Hansen. 1996. Language accent classification in American English. *Speech Communication*, 18(4):353–367.

Mohamad Hasan Bahari, Rahim Saeidi, Hugo Van Hamme, and David Van Leeuwen. 2013. Accent Recognition Using I-vector , Gaussian Mean Supervector and Gaussian Posterior probability Supervector for Spontaneous Telephone Speech. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 7344–7348.

Utpal Bhattacharjee and Kshirod Sarmah. 2012. Gmm-ubm based speaker verification in multilingual environments. *International Journal of Computer Science Issues (IJCSI)*, 9(6):373.

W. M. Campbell, D. E. Sturim, D. A. Reynolds, and A. Solomonoff. 2006. Svm based speaker verification using a gmm supervector kernel and nap variability compensation. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, volume 1, pages I–I.

Nancy F. Chen, Sharon W. Tam, Wade Shen, and Joseph P. Campbell. 2014. Characterizing phonetic transformations and acoustic differences across English dialects. *IEEE Transactions on Audio, Speech and Language Processing*, 22(1):110–124.

Jian Cheng, Nikhil Bojja, and Xin Chen. 2013. Automatic accent quantification of indian speakers of english. In *INTERSPEECH*.

P. Day and A. K. Nandi. 2007. Robust text-independent speaker verification using genetic programming. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(1):285–295.

George R. Doddington. 2001. Speaker recognition based on idiolectal differences between speakers. In *INTERSPEECH*.

Stephen J. Eady and William E. Cooper. 1986. Speech intonation and focus location in matched statements and questions. *The Journal of the Acoustical Society of America*, 80(2):402–415.

Mireia Farrús, Michael Wagner, Daniel Erro, and Javier Hernando. 2010. Automatic speaker recognition as a measurement of voice imitation and conversion. *International Journal of Speech, Language and the Law*, 17(1):119–142.

Horacio Franco, Leonardo Neumeyer, Vassilios Digalakis, and Orith Ronen. 2000. Combination of machine scores for automatic grading of pronunciation quality. *Speech Communication*, 30(2):121–130.

Shahram Ghorbani, John H L Hansen, Robust Speech, and Systems Crss. 2018. Leveraging native language information for improved accented speech recognition. (September):2449–2453.

R. Krishnan G.Radha Krishna and Vinay Kumar Mittal. 2018. Native Language Identification from South Indian English Speech. In *Workshop on Machine Learning in Speech and Language Processing, September 7th, 2018*.

Calbert Graham and Brechtje Post. 2018. Second language acquisition of intonation: Peak alignment in American English. *Journal of Phonetics*, 66:1–14.

Radha Krishna Guntur, R Krishnan, and V.K. Mittal. 2018. Prosodic Analysis of Non-Native South Indian English Speech. In *Proc. The 6th Intl. Workshop on Spoken Language Technologies for Under-Resourced Languages*, pages 71–75.

A. Hanani, M. J. Russell, and M. J. Carey. 2013. Human and computer recognition of regional accents and ethnic groups from British English speech. *Computer Speech and Language*, 27(1):59–74.

Mary P. Harper and Michael Maxwell. 2008. *Spoken Language Characterization*, pages 797–810. Springer Berlin Heidelberg, Berlin, Heidelberg.

Hynek Hermansky. 1990. Perceptual linear predictive (plp) analysis of speech. *The Journal of the Acoustical Society of America*, 87(4):1738–1752.

Florian Hönig, Anton Batliner, and Elmar Nöth. 2012. Automatic Assessment of Non-Native Prosody An-notation, Modelling and Evaluation. *Proceedings of the International Symposium on Automatic Detection of Errors in Pronunciation Training (IS ADEPT)*, pages 21–30.

A. Ikeno and J.H.L. Hansen. 2006. Perceptual Recognition Cues in Native English Accent Variation: "Listener Accent, Perceived Accent, and Comprehension". *2006 IEEE International Conference on Acoustics Speed and Signal Processing Proceedings*, 1:I–401–I–404.

Tomi Kinnunen and Haizhou Li. 2010. An overview of text-independent speaker recognition: From features to supervectors. *Speech Communication*, 52(1):12 – 40.

Jacques Koreman. 2018. Category similarity in multilingual pronunciation training. In *Proc. Interspeech 2018*, pages 2578–2582.

G. Radha Krishna and R. Krishnan. 2014. Influence of mother tongue on english accent. In *Proceedings of the 11th International Conference on Natural Language Processing*, pages 63–67, Goa, India. NLP Association of India.

G.Radha Krishna, R.Krishnan, and V.K.Mittal. 2019. An automated system for regional nativity identification of indian speakers from english speech. In

16th IEEE India Council International Conference INDICON 2019 (Accepted).

Haizhou Li, Bin Ma, and Kong Aik Lee. 2013. Spoken language recognition: From fundamentals to practice. *Proceedings of the IEEE*, 101(5):1136–1159.

Kun Li, Xixin Wu, and Helen Meng. 2017. Intonation classification for L2 English speech using multi-distribution deep neural networks. *Computer Speech and Language*, 43:18–33.

Jesús Antonio Villalba López. 2014. Advances on Speaker Recognition in non Collaborative Environments. page 311.

Iker Luengo, Eva Navas, Iñaki Sainz, Ibon Saratxaga, Jon Sanchez, Igor Odriozola, and Inma Hernaez. 2008. Text independent speaker identification in multilingual environments. In *LREC 2008*.

SHERVIN MALMASI and MARK DRAS. 2017. Multilingual native language identification. *Natural Language Engineering*, 23(2):163215.

Leena Mary and B. Yegnanarayana. 2008. Extraction and representation of prosodic features for language and speaker recognition. *Speech Communication*, 50(10):782–796.

V. K. Mittal and B. Yegnanarayana. 2013. Effect of glottal dynamics in the production of shouted speech. *The Journal of the Acoustical Society of America*, 133(5):3050–3061.

Vinay Kumar Mittal and B Yegnanarayana. 2015. Study of characteristics of aperiodicity in noh voices. *The Journal of the Acoustical Society of America*, 137(6):3411–3421.

Vinay Kumar Mittal, B Yegnanarayana, and Peri Bhaskararao. 2014. Study of the effects of vocal tract constriction on glottal vibration. *The Journal of the Acoustical Society of America*, 136(4):1932–1941.

Vinay Kumar Mittal and Bayya Yegnanarayana. 2014. Significance of aperiodicity in the pitch perception of expressive voices. In *INTERSPEECH*.

Julia Monnin and Hélène Loevenbruck. 2010. Language-specific influence on phoneme development: French and drehu data. In *INTERSPEECH*.

N. Moustroufas and V. Digalakis. 2007. Automatic pronunciation evaluation of foreign speakers using unknown text. *Computer Speech and Language*, 21(1):219–230.

Leonardo Neumeyer, Horacio Franco, Vassilios Digalakis, and Mitchel Weintraub. 2000. Automatic scoring of pronunciation quality. *Speech Communication*, 30(2):83–93.

Sergiu Nisioi. 2015. Feature analysis for native language identification. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9041:644–657.

Mohamed Kamal Omar and Jason Pelecanos. 2010. A novel approach to detecting non-native speakers and their native language. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pages 4398–4401.

Yao Qian, Keelan Evanini, Xinhao Wang, David Suendermann-Oeft, Robert A. Pugh, Patrick L. Lange, Hillary R. Molloy, and Frank K. Soong. 2017. Improving sub-phone modeling for better native language identification with non-native english speech. In *INTERSPEECH*.

D. A. Reynolds and R. C. Rose. 1995. Robust text-independent speaker identification using gaussian mixture speaker models. *IEEE Transactions on Speech and Audio Processing*, 3(1):72–83.

Shambhu Nath Saha and Shyamal Kr. Das Mandal. 2015. Study of acoustic correlates of english lexical stress produced by native (l1) bengali speakers compared to native (l1) english speakers. In *INTERSPEECH*.

Björn Schuller, Stefan Steidl, Anton Batliner, Felix Burkhardt, Laurence Devillers, Christian Müller, and Shrikanth Narayanan. 2013. Paralinguistics in speech and language - State-of-the-art and the challenge. *Computer Speech and Language*, 27(1):4–39.

E. Shriberg, L. Ferrer, S. Kajarekar, A. Venkataraman, and A. Stolcke. 2005. Modeling prosodic feature sequences for speaker recognition. *Speech Communication*, 46(3-4):455–472.

Joseph Tepperman and Shrikanth Narayanan. 2008. Better nonnative intonation scores through prosodic theory. In *INTERSPEECH*.

Nigel G Ward, Nigelward@acm Org, Paola Gallardo, and Amanda Stent. 2017. Non-Native Differences in Prosodic-Construction Use. *Dialogue & Discourse*, 8(1):1–30.

Tingyao Wu, Jacques Duchateau, Jean Pierre Martens, and Dirk Van Compernolle. 2010. Feature subset selection for improved native accent identification. *Speech Communication*, 52(2):83–98.

Marcos Zampieri, Alina Maria Ciobanu, and Liviu P. Dinu. 2017. Native Language Identification on Text and Speech. 2013:398–404.

Nengheng Zheng, Tan Lee, and P. C. Ching. 2007. Integration of complementary acoustic features for speaker recognition. *IEEE Signal Processing Letters*, 14(3):181–184.