Spectral Feature Based Kannada Dialect Classification from Stop Consonants

Nagaratna B. Chittaragi[1,2](✉), Pradyoth Hegde[1], Siva Krishna P. Mothukuri[1], and Shashidhar G. Koolagudi[1]

[1] Department of Computer Science and Engineering,
National Institute of Technology Karnataka, Surathkal, Karnataka, India
nbchittaragi@gmail.com, pradyothhegde@gmail.com, msivakrish@gmail.com, koolagudi@nitk.edu.in

[2] Department of Information Science and Engineering,
Siddaganga Institute of Technology, Tumkur, Karnataka, India

Abstract. This study focuses on the investigation of the significance of stop consonants in view of the classification of Kannada dialects. Majority of the studies proposed have shown the existence of evidential differences in the pronunciation of vowels across dialects. However, consonant based studies on dialect processing are found to be comparatively lesser. In this work, eight stop consonants are used for characterization of five Kannada dialects. Acoustic characteristics such as cepstral coefficients, formant frequencies, spectral flux, and rolloff features are explored from spectral analysis of stops. The consonant dataset is derived from standard Kannada dialect dataset consisting of 2417 consonants obtained from 16 native speakers from each dialect. Support vector machine (SVM) and decision tree-based extreme gradient boosting (XGB) ensemble classification methods are employed for automatic recognition of Kannada dialects. The research findings show that the stops existing for shorter duration also convey dialectal linguistic cues. Combination of spectral properties has contributed to the identification of distinct dialect-specific information across Kannada dialects.

Keywords: Kannada dialect classification · Stop consonants · Spectral features · SVM · XGB

1 Introduction

Automatic identification of dialects may be considered as a prominent research area in the speech community due to its implications to automatic speech recognition (ASR) systems. The primary task of the dialect identification is to recognize the speaker's regional variety of the language spoken. Due to overlaps in vocabulary and phonetic similarities among the dialects, dialect identification problem is considered to be more challenging than language identification [1].

The present paper explores the identification of five Kannada dialects from the stop consonant utterances. The Kannada language is an official language

B. Deka et al. (Eds.): PReMI 2019, LNCS 11941, pp. 82–90, 2019.
https://doi.org/10.1007/978-3-030-34869-4_10

spoken in Karnataka state. Very few dialect related studies are found in literature w.r.t. Kannada language. Stop consonants represent one of the broad categories of phones. The production of a stop involves a complete closure of the oral cavity followed by the release in the form of noise burst. The stop consonants are differentiated from each other in terms of the manner of articulation (whether voiced and aspirated) and the place of articulation. Because of their short duration, the classification of stops is a challenging problem.

Literature gives comparatively fewer works available for consonant-based dialect processing.

A study is proposed with the investigation of the existence of dialectal information in the bursts during the production of the stop consonants of two Greek dialects, namely, standard Greek and Cypriot [15]. A study has analyzed the pre-nasalization differences during the production of the voiced stop consonants among two dialects of Greek. The acoustic cues of voice onset time (VOT) and F0 are captured to represent the laryngeal contrast of Korean stops between Chonnam Korean and Seoul Korean dialects. These acoustic features have effectively modelled the cross-dialectal variations among the stops [7]. Cretan and Thessalonikan dialects of Modern Greek are considered and measured amplitude and duration features. This model has suggested the more complex and interactive influences of dialects, gender, and stress in the realization of pre-nasality in the voiced stops [10]. Few systems can be found for addressing dialect processing of Kannada language. An author has considered only two dialects of Kannada earlier and used only pronunciation variations among vowels. They used acoustic features for their analysis [13].

However, very few systems have been found for analysis of Kannada consonants from the dialect perspectives. A study has included the analysis of dialectal variations among four primitive dialects for the task of speech recognition. However, this has not included any analysis w.r.t. stop consonants [9]. Research findings with consonant-based Kannada dialect processing are found to be limited because of the following reasons. 1. Availability of Standard dialect dataset for Kannada language. 2. Accurate segmentation of stops is a difficult task due to the duration of the stops. 3. Plosives comprise the stops and affricates and are considered particularly challenging to recognize because of their highly dynamic characteristics differs with vowels. And this has motivated to consider this problem for dialectal analysis.

In this study, eight un-aspirated unvoiced and voiced consonants namely, /p/, /b/, /T/, /t/, /D/, /d/, /k/, and /g/ are used for classification of Kannada dialects. These eight consonants are commonly known as plosive sounds, as they are produced due to the constriction occurred at different regions in the mouth. Among these, /k/ & /g/ are velar, /p/ & /b/ are labial, /t/ & /d/ are dental, and /T/ & /D/ are retroflex consonants. Plosives of Kannada have typically observed with silent period in the closure phase, as closure duration of /k/, /t/, /T/ and /p/ are longer when compared to /g/, /d/, /D/ and /b/. Also, voiced plosives show the voicing bar at the lower frequencies. A stronger vertical spike is observed showing release burst for /k/, /t/ and /p/.

Table 1. Kannada dialect dataset

	CENK	CSTK	HYDK	MUBK	STHK
No. of speakers	30	34	37	26	29
Gender (Male+Female)	18+12	19+15	25+12	12+14	16+13
Duration (in min.)	112	132	120	130	128

Occurrences of both palatal stops in natural communication are comparatively rare in Kannada language. Hence, in this work, four types of stops produced at the velar, dental, retroflex, and labial consonants are analyzed individually to see their significances across five dialects. Spectral characteristics representing features such as the first three formant frequencies, Mel frequency cepstral coefficients (MFCCs), spectral flux, centroid, rolloff, and entropy are extracted from stops. Dialect classification systems are developed with SVM and XGB ensemble techniques.

The present paper is organized as follows. Consonant dataset details are provided in Sect. 2. Details of extracted spectral features and employed classification models are discussed in Sect. 3. Information regarding the experiments carried out, results and analysis are given in Sect. 4. Conclusions of the current work are presented in Sect. 5 along with future directions.

2 Consonant Based Dialect Dataset

The consonant dataset used in this work is derived from spontaneous Kannada dialect dataset [5]. It consists of five distinct Kannada dialects spoken across Karnataka, namely: Central Karnataka Region (CENK), Coastal Karnataka Region (CSTK), Hyderabad Karnataka Region (HYDK), Mumbai Karnataka Region (MUBK), and Southern Karnataka Region (STHK). Text-independent spontaneous speech is recorded from native dialect speakers. The age group of speakers lies between 21–72 years. Majority of the speakers are from rural areas, and they are not moved to any other places for a long time. An interview style is followed to make speakers talk continuously in a reasonably quiet environment. Detailed information regarding Kannada dialect dataset is presented in Table 1.

It is observed that the duration of plosive sounds is concise; hence, segmentation and extraction of significant dialect-specific features from them is a tedious task. Hence, in this work, all consonants collected are in terms of monosyllables (CV units), and plosive sounds are combined with vowels /a/, /u/, /i/ or /o/. Since co-articulation between them is comparatively less [11]. The /CV/ tokens are identified in spoken utterances manually using Praat tool. The onset of stop burst and offset of the vowels are identified through simultaneous inspections of both waveform and the spectrogram [2]. /CV/ syllables are segmented by detecting burst onset, /CV/ transition and complete vowel utterances. Majority of the /CV/ units are chosen from the word-initial position. A plosive dataset considered in this study consists of total 2417 stops extracted from 16 (9 Female + 7

Male) speakers from each dialect. The total number of consonant clips available for each dialect are as follows, CENK-455, CSTK-478, HYDK-484, MUBK-501, STHK-499.

3 Spectral Feature Extraction and Classification Models

In this paper, acoustic characteristics those significantly differentiates stops consonants of five dialects of Kannada are extracted. Features such as formants, standard MFCCs, spectral flux, centroid, rolloff, and entropy are explored to represent spectral behavior. These features try to capture vocal tract variations among different consonants across Kannada dialects.

3.1 Spectral Feature Extraction

Formants: Due to the oral cavity is closed during the constriction at the specific place at vocal tract, formants may not be available. However, closure is followed by the release of noise burst, due to which front cavity is excited by a sudden reduction in downstream. This, in turn, leads in shifting of formants either upwards or downward along with amplitude. This is depending on the place of constriction of the consonant and the following vowel. Three formant frequencies play a significant role in identifying different vowels. Similarly, vowels, along with stops, can also encapsulate variations occur during the pronunciation variations [14]. LPC based McCandless formant tracking algorithm is employed to extract the three formant frequencies from plosives with a 10 ms overlapped 20 ms frame [12]. LPC is a widely used method for formant extraction due to its compact and accurate computation. Figure 1 presents the utterance of /k/ consonant manually segmented from the word “Kannada” from five male speakers of five Kannada dialects.

MFCCs: Spectral changes between the five Kannada dialects are captured using MFCC features. These features have proven to capture the vocal tract variations of speech signal successfully. The coefficients extracted resembles the human auditory system. 13 MFCCs are extracted from a speech signal using block processing approach from a 20 ms frame with a shift of 10 ms.

Spectral Flux: Timbre is the speaker-specific feature of the sound unit that helps to compare the similarity of two speech utterances. The spectral flux usually corresponds to a perceptual roughness of sound. In this work, flux feature is computed and used to measure the spectral changes existing between two successive frames. It is computed by extracting the power spectra of one frame against the same of the previous one [8].

$$Fl_{(i,i-1)} = \sum_{k=1}^{Wf_L} (EN_i(k)) - (EN_{i-1}(k))^2 \tag{1}$$

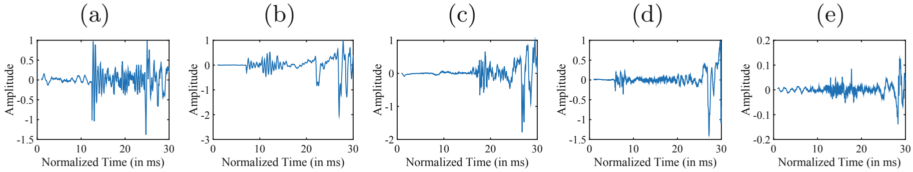

Fig. 1. Utterance of /k/ manually segmented from word "Kannada" five male speakers of five Kannada dialects, (a) CENK, (b) CSTK, (c) HYDK, (d) MUBK, (e) STHK

where $EN_i(k) = \frac{X_i(k)}{\sum_{l=1}^{Wf_L} X_i(l)}$, here $EN_i(k)$ is the k^{th} normalized DFT coefficient at the i^{th} frame, Wf_L is the frame size.

Spectral Entropy: Spectral entropy of a signal measures the distribution of spectral power. Spectral entropy is used to capture the abrupt changes within the energy levels of consonants. While computing spectral entropy of a frame, corresponding spectrum is divided into L sub-bands (bins). The energy E_f of the f^{th} sub-band, for f = 0, . . ., L-1 is calculated using Eq. (2). Then, energies of all bins are normalized by dividing with the spectral energy of the whole frame, i.e., $ef = \frac{E_f}{\sum_{f=0}^{L-1} E_f}$, the entropy of each normalized energy value is calculated using the Eq. (3)

$$E(i) = \frac{1}{Wf_L} \sum_{k=1}^{w_L} |x_i(k)|^2 \tag{2}$$

$$H = -\sum_{f=0}^{L-1} ef.\log_2(ef) \tag{3}$$

In this work, the value of L is set to 10 indicating that each frame is divided into 10 bins.

Spectral Roll-Off: Spectral rolloff feature is treated as a spectral shape descriptor of an audio signal and it is used to differentiate voiced and unvoiced sounds. This feature is defined as the frequency below which a certain percentage (generally 90%–95%) of the magnitude distribution of the spectrum is concentrated.

3.2 Classification Models

In this work, standard SVM and multiple classifiers based XGB algorithms are used for the implementation of dialect recognition systems from stops. The SVM classification method is employed to capture dialect-specific cues. It is designed with the one-versus-rest approach to handle the 5-class pattern classification problem. Radial basis function (RBF) kernel function is used for separating hyperplane with the maximal margin in a high dimensional feature space [3]. Apart from this, nowadays, ensemble algorithms are gaining popularity. These

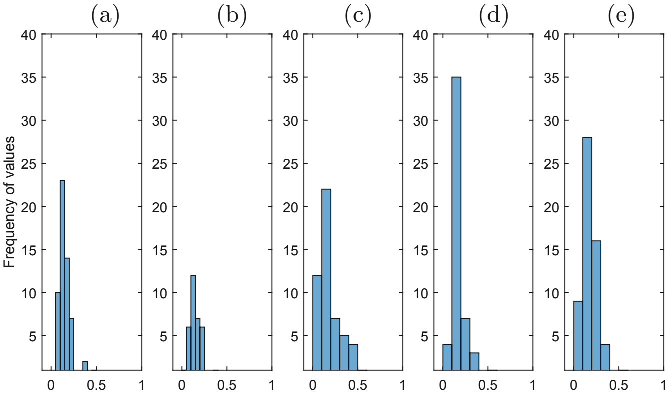

Fig. 2. Spectral rolloff variations among five Kannada dialects, (a)-CENK, (b)-CSTK, (c)-HYDK, (d)-MUBK, (e)-STHK

are powerful prediction and classification techniques in enhancing the performance with a combination of multiple classifiers over single classification methods. However, use of these methods for speech processing in specific for dialect identification is rarely found [6]. Hence, XGB method is used for classifying five dialects. Among ensembles, the most commonly used gradient boosting involves three significant steps: The first step is the selection of a suitable loss function; here, multi-class *logloss* is used as the problem addressed is a classification problem. Next step is choosing a base classifier, and decision trees are used in this paper, where trees are constructed using the greedy approach. Few parameters such as best split, number of leaf nodes, maximum levels are fine-tuned to produce a better performance. In the third step, trees are added one at a time; a gradient descent procedure is used for minimization of loss during the addition of trees. The XG boost library is used for implementation to handle five dialect classes [4].

4 Experiments and Results

This section provides details of the spectral analysis of stops through features explored from stops from five dialects along with complete details of experiments and results obtained.

From Fig. 1, it can be noticed that there are variations in the pronunciation of consonant /k/ across five dialects. Varying length and energy of the burst regions can be seen among five dialects. Spectral rolloff is generally treated as a spectral shape descriptor of an audio signal and it is usually used for discrimination of voiced and unvoiced sounds. In this study, the histograms are plotted for the spectral rolloff feature and are presented in Fig. 2. The rolloff parameters are extracted from /k/ stops spoken from five dialect speakers. From these histograms, it has been observed that the spectral rolloff value distributions

Table 2. Average dialect recognition performance from **consonants** level utterances using SVM and XGB methods (Accuracies in %)

Sl. no.	Features extracted	Kannada dialects	
		SVM	XGB
1	MFCCs	73.33	78.00
2	Spectral flux, rolloff, centroid, F1 and F2	59.37	68.76
3	Spectral flux, rolloff, centroid, F1 and F2 + MFCCs	77.65	**78.33**

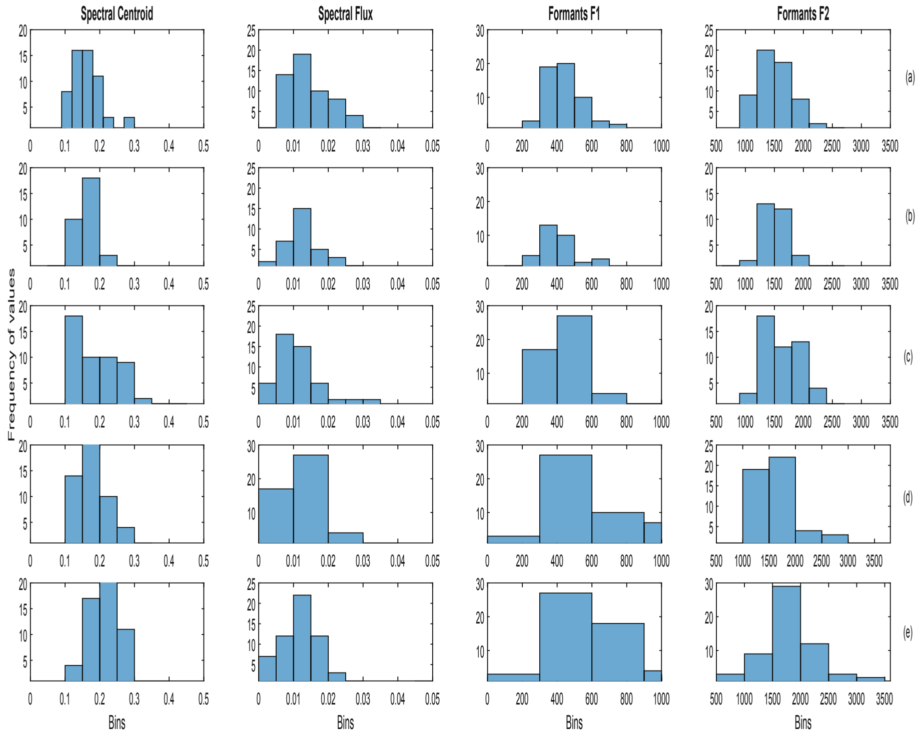

Fig. 3. Histograms of four spectral features, drawn for utterance of /k/ from five dialect speakers. (a) CENK, (b) CSTK, (c). HYDK, (d) MUBK, (e) STHK

are comparatively lower with CSTK and CENK dialects over three remaining dialects. It is also observed that the dialect MUBK and STHK dialect is with higher values of the spectral rolloff sequence. Besides, the variation is more intense for these two dialects. Whereas, CSTK and CENK dialects are noticed with lower values of spectral rolloff.

In order to show the differences across dialects with consonant /k/, centroid, flux, and two formants (F1 and F2) features are considered. The same above mentioned histogram is used. Figure 3 is drawn to show the spectral variations across dialects. Formants F1 and F2 are seen to be the distinguishing features among dialects with /k/ stops. However, the spectral centroid is seen with lesser variations. Even spectral flux feature considered is also contributing to several differences among dialects. Based on the analysis of the features as mentioned above in the characterization of dialects, several experiments are conducted by using these features for classification of dialects. In this work, single and ensemble SVM classification methods are employed for the development of Kannada dialect classification systems. Average dialect classification performance obtained from consonant utterances are presented in Table 2.

From the results obtained from the Table 2, it is observed that MFCCs are seen to be powerful features in classification of dialects from very shorter utterances such as stop consonants. Even spectral features are also captured the dialectal cues from stops. The combination of features has demonstrated an accuracy of 78.33%, which is slightly higher than MFCC features alone. However, from these analyses, it is noticed that dialect specific evidence are present even at consonant level utterances. Spectral attributes can effectively model the dialectal variations of stop consonants.

5 Conclusions

The present paper has proposed a Kannada dialect identification system from the stop consonants. For this purpose, spectral features are extracted from the shorter stop consonants to capture dialect-specific features. Stop consonant dataset is used in this is created from Kannada dialect dataset. SVM and XGB classification models are employed for automatic recognition of dialects. Combination of spectral features has demonstrated the better classification performance over MFCC features alone. A highest of 78.33% of dialect recognition performance is achieved with the use of stop consonants. However, in this paper, consonants considered are associated with vowels sounds. In the future, consonants alone can be used for classification of dialects. Dialect-specific features can be explored from the burst regions of the consonants as they consist of more relevant dialectal information. Apart from spectral analysis, excitation source features can also enhance dialect recognition performance.

Acknowledgment. This work is supported by DST-GOI (Department of Science and Technology, Government of India) sponsored project entitled *Characterization and Identification of Dialects in the Kannada Language.*

References

1. Biadsy, F.: Automatic dialect and accent recognition and its application to speech recognition. Ph.D. thesis, Columbia University (2011)
2. Boersma, P.: Praat, a system for doing phonetics by computer. Glot Int. **5**(10), 341–345 (2002)
3. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. (TIST) **2**(3), 27 (2011)

4. Chen, T., Guestrin, C.: XGBoost: a scalable tree boosting system. In: Twenty second International Conference on Knowledge Discovery and Data Mining, pp. 785–794. ACM (2016)
5. Chittaragi, N.B., Limaye, A., Chandana, N.T., Annappa, B., Koolagudi, S.G.: Automatic text-independent kannada dialect identification system. In: Satapathy, S.C., Bhateja, V., Somanah, R., Yang, X.-S., Senkerik, R. (eds.) Information Systems Design and Intelligent Applications. AISC, vol. 863, pp. 79–87. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-3338-5_8
6. Chittaragi, N.B., Prakash, A., Koolagudi, S.G.: Dialect identification using spectral and prosodic features on single and ensemble classifiers. Arab. J. Sci. Eng. **43**(8), 4289–4302 (2018)
7. Choi, H.: Acoustic cues for the Korean stop contrast-Dialectal variation. Citeseer (2013)
8. Giannakopoulos, T., Pikrakis, A.: Introduction to Audio Analysis: A MATLAB Approach. Academic Press, Cambridge (2014)
9. Hemakumar, G., Punithavalli, M., Thippeswamy, K.: Speech recognition system for different Kannada dialects. Int. J. Sci. Res. Comput. Sci. Eng. Inform. Technol. **2**(5), 180–188 (2017)
10. Jong Kong, E., Syrika, A., Edwards, J.R.: Voiced stop prenasalization in two dialects of greek. J. Acoust. Soc. Am. **132**(5), 3439–3452 (2012)
11. Kalaiah, M.K., Bhat, J.S.: Effect of vowel context on the recognition of initial consonants in kannada. J. Audiol. Otol. **21**(3), 146 (2017)
12. McCandless, S.: An algorithm for automatic formant extraction using linear prediction spectra. IEEE Trans. Acoust. Speech Signal Process. **22**(2), 135–141 (1974)
13. Nagesha, K.S., Kumar, G.H.: Acoustic-phonetic analysis of Kannada accents. Tata Institute of Fundamental Research, Mumbai (2010)
14. Reetz, H., Jongman, A.: Phonetics Transcription, Production, Aoustics and Perception. Wiley Blackwell, Hoboken (2009)
15. Themistocleous, C.: The bursts of stops can convey dialectal information. J. Acoust. Soc. Am. **140**(4), EL334–EL339 (2016)