Chapter 2
Language Identification—A Brief Review

Abstract This chapter provides compendious reviews about both the explicit and implicit LID systems present in the literature. Existing works related to language identification in Indian context are briefly discussed. The related works about the excitation source features are also presented here. Various speech features and models proposed in the context of language identification are briefly reviewed in this chapter. The motivation for the present work from the existing literature is briefly discussed.

Keywords Prior works on explicit language identification · Prior works on implicit language identification · Prior works on excitation source features · Motivation for using source features for language identification

2.1 Prior Works on Explicit Language Identification System

In 1974, Dodington and Leonard [1], Leonard [2] have explored frequency of occurrences of certain reference sound units in different languages. The average LID accuracy of 64 % and 80 % have been achieved using five and seven languages, respectively.

In 1977, House and Neuberg [3] conducted LID studies on manually phonetic transcribed data. The language related information has been extracted from a broad phonetic transcription instead of using acoustic features extracted from speech signal. Speech signal has been considered as a sequence of symbols chosen from a set. The elements of the set are defined as follows: stop consonant, fricative consonant, vowel and silence. Language identification experiment has been carried out on eight languages. In this work, Hidden Markov Model (HMM) has been trained using broad phonetic labelled data derived from phonetic transcription. This work had shown perfect discrimination of eight languages and demonstrated that excellent language identification can be achieved by exploiting phonotactic information.

In 1980, Li and Edwars [4] developed automatic LID system based on automatic acoustic-phonetic segmentation of speech. By using six different acoustic-phonetic classes, automatic LID system has been developed using five languages. These six acoustic-phonetic classes are (i) syllable nuclei, (ii) non-vowel sonorants, (iii) vocal

K.S. Rao and D. Nandi, *Language Identification Using Excitation Source Features*, SpringerBriefs in Speech Technology,
DOI 10.1007/978-3-319-17725-0_2

murmur, (iv) voiced frication, (v) voiceless frication and (vi) silence and low energy segments. Hidden Markov Models (HMM) have been used for developing language models. Recognition accuracy of 80 % has been achieved with this approach.

In 1993 and 1994, Lamel and Gauvain [5, 6] conducted cross-lingual experiments by exploring phone recognition for French and English languages. A set of 35 phones were used to represent the French language corpus and a set of 46 phones were used to represent the English language data. Three-state left-to-right continuous density HMM with Gaussian mixture model (GMM) observation density has been used to build the phone models. It has been observed that, the French language is easier to recognize at the phone level but, harder to recognize at the lexical level due to the larger number of homophones.

In 1994, Muthusamy et al. [7] have proposed a perceptual benchmark for language identification task. Perceptual studies with listeners from different language backgrounds have been conducted. The experiments have been conducted on ten languages from OGI-MLTS database. The results obtained from the subjects reported as the benchmark for evaluating the LID performances obtained from automatic LID systems. The experimental analysis showed that, the duration of utterances, familiarity of languages and the number of known languages are the important factors to recognize a language. The comparison between the subjective analysis and machine performance concluded that, increased exposure to each language and longer training sessions contribute to improved classification performance. Therefore, to develop the speech recognizer for any language, the primary requirement is large amount of segmented and labelled speech corpus.

In 1994, Berkling et al. [8] have analyzed phoneme based features for language recognition. They have performed the LID study on three languages: English, Japanese and German from OGI-MLTS speech corpus. A superset of phonemes for the three languages has been considered. The phonemes which can provide the best discrimination between language pairs have used to build the superset. The experimental analysis drawn the conclusion that, to develop a LID system with large number of languages, it may be useful to reduce the number of features despite a small loss in LID accuracy.

In 1994, Tucker et al. [9] have conducted LID experiments with the languages belong to same language family. Sub-word models for English, Dutch and Norwegian languages have been developed for carrying out the LID study. Two types of language models: language independent and language-specific models have been developed in this study. Three techniques namely, (i) the acoustic differences between the phonemes of each language, (ii) the relative frequencies of phonemes of each language and (iii) the combination of previous two sources have been explored for classifying the languages. The third technique provides average LID accuracy of 90 % for three languages.

In 1994, Zissman and Singer [10] have carried out a comparative study using four approaches: (i) Gaussian mixture model based classification, (ii) phoneme recognition followed by language modeling (PRLM), (iii) parallel PRLM (PRLM-P) and (iv) language-dependent parallel phoneme recognition (PPR). The OGI-MLTS corpus has been used to evaluate the performances of the four LID approaches. The LID

study showed that, best performance is obtained with PRLM-P system, which does not require labelled speech corpus for developing language models.

In 1995, Kadambe and Hieronymus [11] have developed LID systems using phonological and lexical models to distinguish the languages. The LID study has been carried out on four languages: English, German, Mandarin and Spanish from OGI-MLTS speech corpus. Identification accuracy of 88 % has been achieved with four languages. It has been observed that, English and Spanish languages are distinguishable by their lexical information. This study concludes that, the language-specific information can also be captured by analyzing the higher level linguistic knowledge.

In 1995, Yan and Bernard [12] have developed language-dependent phone recognition systems for language discrimination task. Six languages (English, German, Hindi, Japanese, Mandarin and Spanish) from OGI-MLTS corpus have been used for LID study. Continuous HMMs are used to build the language-dependent phone recognizers. Acoustic and duration models are exploited for developing LID system. Forward and backward bigram based language models are proposed. A neural network based approach has been proposed for combining the evidences obtained from the above mentioned acoustic, language and duration models.

In 1997, Navratil and Zhulke [13] have proposed two approaches to build language models: (i) modified bigrams with a context mapping matrix and (ii) language models based on binary decision trees. To build the binary decision tree two approaches are proposed. These two approaches are, (i) building the whole tree for each class and (ii) adapting from a universal background model (UBM). Both the models are incorporated in a phonetic language identifier with a double bigram decoding architecture. The LID study has been carried out on NIST'95 language database.

In 1997, Hazen and Zue [14] have developed automatic LID system utilizing the phonotactic, acoustic-phonetic and prosodic information within a unified probabilistic framework. The evidences obtained from three different sources are combined to improve the LID accuracy. Experimental results showed that, the phonotactic information present in the speech utterances is the most useful information for language discrimination task. It has been observed that, acoustic-phonetic and prosodic information can also be useful for increasing the system's accuracy, especially when the short duration utterances are used for evaluation.

In 2001, K. Kirchhoff and S. Parandekar [15] have developed LID systems based on n-gram models of parallel streams of phonetic features and sparse statistical dependencies between these streams. The LID study has been conducted on OGI-MLTS database. It has been shown that, the proposed feature-based approach outperforms phone-based system. They have also reported that, proposed approach shows significantly better identification accuracy using test utterances of very short duration ($\leq$3 s). In future, data-driven measures for predicting optimal cross-stream dependencies, as well as different schemes for score integration can be explored.

In 2001, Gleason and Zissman [16] have demonstrated two methods to enhance the accuracy of parallel PRLM (PPRLM) system. They have explored Composite background (CBG) modeling technique, which allows us to identify target language in an environment where labelled training data is unavailable or limited.

In 2003, V. Ramasubramanian et al. [17] have shown the theoretical equivalence of parallel sub-word recognition (PSWR) and Ergodic-HMM (E-HMM) based LID. In this work, the sub-word recognizer (SWR) at the front-end represents the states and the language model (LM) of each language at the back-end represents the state-transition of E-HMM in that language. The proposed equivalence unifies two distinct approaches of langauge identification: parallel phone (sub-word) recognition and E-HMM based approaches. This LID study has been carried out on 6 languages from OGI-MLTS database. The performance of E-HMM based system is superior compared to GMM, which indicates the effectiveness of the E-HMM based approaches.

In 2004, J. Gauvain et al. [18], Shen et al. [19] proposed a novel method using phone lattices for developing automatic LID system. The use of phone lattices both in training and testing significantly improves the accuracy of a LID system based on phonotactics. Decoding is done by maximizing the expectation of the phonotactic likelihood for each language. Neural network has been used to combine the scores of multiple phone recognizers for improving the recognition accuracy. NIST 2003 corpus is used for carrying out the study.

In 2007, Li et al. [20] have proposed a novel approach for spoken language identification task based on vector space modeling (VSM). The hypothesis is that, the overall characteristics of all languages can be covered by a universal set of acoustic units, which can be characterized by the acoustic segment models (ASMs). The ASM framework further extended to language independent phone models for LID task by introducing an unsupervised learning procedure to circumvent the need for phonetic transcription. The spoken utterance has been converted to a feature vector with its attributes representing the co-occurrence statistics of the acoustic units. Then a vector space classifier has been built for language identification. The proposed framework has been evaluated on NIST 1996 and 2003 LRE databases.

In 2008, Sim and Li [21] have proposed a new approach for building a parallel phone recognition followed by language model (PPRLM) system. A PPRLM system comprises multiple parallel sub-systems, where each sub-system employs a phone recognizer with a different phone set for a particular language. This method aims at improving the acoustic diversification among its parallel sub-systems by using multiple acoustic models. The acoustic models are trained on the same speech data with the same phone set but using different model structures and training paradigms. They have examined the use of various structured precision (inverse covariance) matrix modeling techniques as well as the maximum likelihood and maximum mutual information training paradigms to produce complementary acoustic models. The results show that, acoustic diversification, which requires only one set of phonetically transcribed speech data, yields similar performance improvements compared to phonetic diversification. In addition, further improvements were obtained by combining both diversification factors. The proposed approach has been evaluated on NIST 2003 and 2005 LRE databases.

In 2008, Tong et al. [22] have proposed a target-oriented phone tokenizers (TOPT), each having a subset of phones that have high discriminative ability for a target language. Two phone selection methods are proposed to derive such phone subsets from a phone recognizer. It has been shown that, the TOPTs derived from a universal

phone recognizer (UPR) outperform those derived from language specific phone recognizers. The TOPT front-end derived from a UPR also consistently outperforms the UPR front-end without involving additional acoustic modeling. The proposed method has been evaluated on NIST 1996, 2003 and 2007 LRE databases.

In 2012, Botha and Barnard [23] used n-gram statistics as features for LID study. A comparative study has been carried out using different classifiers such as, support vector machines (SVMs), naive Bayesian and difference-in-frequency classifiers. The work has been carried out by varying the values of n. Experimental results conclude that, the SVM classifier outperforms other classifiers.

In 2012, Barroso et al. [24] have proposed hybrid approaches to build LID system based on the selection of system elements by several classifiers (Support Vector Machines (SVMs), Multilayer Perceptron classifiers and Discriminant analysis). The LID study has been carried out on three languages: Basque, Spanish and French. The proposed approach improves the system performance.

In 2013, Siniscalchi et al. [25] proposed a novel universal acoustic characterization approach for language recognition. Universal set of fundamental units has been explored, which can be defined across all the languages. This LID study has exploited some speech attributes like manner and place of articulations of sound units to define the universal set of language-specific fundamental units. Summary of the prior works related to explicit LID studies mentioned above is provided in Table 2.1.

Table 2.1 Summary of prior works on *explicit* language identification studies

Sl. no.	Features	Models/ Classification techniques	Number of languages and databases	Remarks	Reference
1.	Broad phonetic transcription (i.e., stop consonant, fricative consonant, vowel and silence)	HMM	8 languages	Phonotactic information is language-specific	[3]
2.	Acoustic-phonetic information	HMM	5 languages	Recognition accuracy of 80 % has been achieved	[4]
3.	PLP coefficients with 56 dimensions	ANN	3 languages from OGI-MLTS database	To develop LID system with large number of languages, it may be useful to reduce the number of features despite a small loss in LID accuracy	[8]

(continued)

Table 2.1 (continued)

Sl. no.	Features	Models/ Classification techniques	Number of languages and databases	Remarks	Reference
4.	Acoustic differences between the phonemes, relative frequency of phonemes and combination of previous two sources of information	Sub-word models were built using HMM	8 languages from EUROM 1 database	90 % LID accuracy is achieved	[9]
5.	MFCC	GMM, PRLM, PRLM-P, PPR	10 languages from OGI-MLTS database	PRLM-P provides best accuracy of 79.2 %	[10]
6.	Phoneme inventory, phonemotactics, syllable structure, lexical and prosodic differences	HMM	4 languages from OGI-MLTS database	88 % accuracy is achieved. Language-specific information can be captured using higher order linguistic knowledge	[11]
7.	Acoustic and duration models	HMM for phoneme recognizer and forward ANN for combining the scores	6 languages from OGI-MLTS database and backward bigram based language model	91.06 % accuracy is achieved for test sample length of 45 s	[12]
8.	Information from a wider phonetic context	Modified bigrams with a context mapping matrix and language models based on binary decision trees	9 languages NIST'95 LRE database	Error rate of 9.4 % is achieved with 45 s test sample duration	[13]
9.	Phonotactic, acoustic-phonetic and prosodic information	Interpolated trigram model and GMM	OGI-MLTS database	Phonotactic information is most useful information for LID task	[14]

(continued)

Table 2.1 (continued)

Sl. no.	Features	Models/ Classification techniques	Number of languages and databases	Remarks	Reference
10.	Phonetic features like, voicing, consonantal place of articulation, manner of articulation, nasality and lip rounding	HMM for phone recognition and *n*-gram language model	OGI-MLTS database	Proposed feature-based approach outperforms phone-based system	[15]
11.	MFCC	Parallel sub-word recognition and Ergodic HMM based LID	6 languages of OGI-MLTS database	The performance of E-HMM based system is superior compared to GMM	[17]
12.	Lexical constraints and phonotactic patterns	PPRLM	NIST 1996, 2003 and 2007 LRE databases	TOPTs derived from UPRs outperform those from language-specific phone recognizers	[22]
13.	*n*-gram statistics as features used for text based LID	SVM, naive Bayesian and difference-in-frequency classifiers	11 South African languages	The SVM classifier outperforms other classifiers and 99.4 % accuracy is achieved	[23]
14.	Morphological features	Hybrid system using SVM, Multilayer Perceptron classifiers and Discriminant analysis	3 languages in Basque context	Hybrid approach works well for under-resourced languages	[24]
15.	Manner and place of articulations of sound units	SVM and maximal figure-of-merit (MFoM)	NIST 2003	Universal set of language-specific fundamental units is proposed	[25]

2.2 Prior Works on Implicit Language Identification System

In 1986, Foil [26] has explored two different approaches to carry out LID study in noisy background. In first approach, language-specific prosodic features are captured by processing pitch and energy contours for LID task. Even though the languages with very similar phoneme sets, the frequency distribution of phonemes often

vary between the languages. In second method, formant vectors are computed only from the voiced segments for each language which is used to discriminate the same phonemes present in different languages. K-means clustering algorithm has been used for formant classification. The conclusion has been drawn from this LID study is that, formant features are better than the prosody features for LID task.

In 1989, Goodman et al. [27] have improved the LID accuracy obtained by Foil [26] in 1986. An important modification has been suggested to training algorithm. The training data has been split into "clean" and "noisy" vectors. K-means clustering algorithm has been used in this experiment. Experiments were also carried out to determine whether pitch information is useful in performing language identification in such noisy conditions or not. The use of syllabic rate as a language discriminative feature has also been investigated.

In 1991, Muthusamy et al. [7] have proposed a phonetic segment-based approach for developing automatic spoken language identification system. The idea was that, the acoustic structure of languages can be estimated by segmenting speech into broad phonetic categories. The language-specific phonetic and prosodic information has been extracted to develop automatic LID system. The LID study has been carried out on American English, Japanese, Mandarin Chinese and Tamil languages. Identification accuracy of 82.3 % has been achieved.

In 1991, Sugiyama [28] has explored linear prediction coefficients (LPCs) and cepstral coefficients (LPCCs) for language recognition. Vector quantization (VQ) of different code book sizes has been proposed for language recognition task. Different distortion measurement techniques like cepstral distance and weighted likelihood ratio have been explored in this LID study. In [9], VQ histogram algorithm has also been proposed for language prediction. Morgan et al. [29] and Zissman [30] have proposed the Gaussian mixture models (GMMs) [31] for language identification study.

In 1994, Itahashi et al. [32] and Shuichi and Liang [33] have developed LID systems based on fundamental frequency and energy contours with the modeling technique based on a piecewise-linear function.

In 1994, K. Li [34] explored spectral features at syllable level to capture the language discriminative information. The syllable nuclei (vowels) are detected automatically. The spectral feature vectors are then computed from the regions near the syllable nuclei instead of computing feature vectors from the whole training data.

In 1999, F. Pellegrino and R. Andre-Obrecht [35] have designed a unsupervised approach based on vowel system modeling. In this work, the language models are developed only using the features extracted from the vowels of each language. Since this detection is unsupervised and language independent, no labelled data is required. GMMs are initialized using an efficient data-driven variant of the LBG algorithm: the LBG-Rissanen algorithm. This LID study are carried out on 5 languages from OGI-MLTS database which provides 79 % recognition accuracy.

In 2005, Rouas et al. [36] have proposed an approach for language identification study based on *rhythmic* modelling. Like phonetics and phonotactics, *rhythm* is also an important feature which can be used for capturing language-specific information. In [36] an algorithm has been proposed to extract the *rhythm* for LID task. They

have used a vowel detection algorithm to segment the *rhythmic* units related to syllables. Several parameters are extracted (consonantal and vowel duration, cluster complexity) and modelled with a Gaussian Mixture. This LID study has been carried out on read speech collected from seven languages.

In 2007, Rouas [37] have developed a LID system based on modelling the prosodic variations. n-gram models were used to model the short-term and long-term language-dependent sequences of labels. The performance of the system is demonstrated by experiments on read speech and evaluated by experiments on spontaneous speech. An experimental study has also been carried out to discriminate the Arabic dialects. It has been shown that the proposed system was able to clearly identify the dialectal areas, leading to the hypothesis that, Arabic dialects have prosodic differences.

In 2010, Sangwan et al. [38] have proposed a language analysis and identification system based on the speech production knowledge. The proposed method automatically extracts key production traits or "hot-spots" which have significant language discriminative capability. At first, the speech utterances were parsed into consonant and vowel clusters. Subsequently, the production traits for each cluster is represented by the corresponding temporal evolution of speech articulatory states. It was hypothesized that, a selection of these production traits are strongly tied to the underlying language, and can be exploited for identifying languages. The LID study has been carried out on 5 closely related languages spoken in India namely, Kannada, Tamil, Telugu, Malayalam, and Marathi. The LID accuracy of 65 % is achieved with this approach. Furthermore, the proposed scheme was also able to identify automatically the key production traits of each language (e.g., dominant vowels, stop-consonants, fricatives etc.).

In 2012, Martnez et al. [39] have proposed an i-vector based prosodic system for language identification system. They have built an automatic language recognition system using the prosody information (*rhythm*, *stress*, and *intonation*) from speech and makes decisions about the language with a generative classifier based on i-Vectors.

In Indian context, J. Balleda et al. [40], have first attempted to identify Indian languages. VQ and 17 dimensional mel-frequency cepstral coefficients (MFCCs) have been explored for language recognition task. Nagarajan [41], have explored different code book methods for LID study. Automated segmentation of speech into syllable like units and parallel syllable like unit recognition have been explored to build *implicit* LID system. Sai Jayaram et al. [42], have proposed trained sub-word unit models without any labelled or segmented data, which are clustered using K-means clustering algorithm. Hidden Markov models (HMM) are used for predicting the language. In 2004, Leena Mary and B. Yegnanarayana have explored the autoassociative neural networks (AANN) for capturing language-specific features for developing LID system [43]. They have also explored prosodic features for capturing the language-specific information [44]. In K.S. Rao et al. [45], have explored spectral features using block processing (20 ms block size), pitch synchronous and glottal closure region (GCR) based approaches for discriminating 27 Indian languages. The language-specific prosodic features have also been explored by V. R. Reddy

et al. [46]. In this work, prosodic features are extracted from syllable, word and sentence levels to capture language-specific information. Jothilakshmi et al. [47], have explored a hierarchical approach for identifying the Indian languages. This method first identifies the language group of a given test utterance and then identifies the particular language inside that group. They have carried out the LID task by using different acoustic features such as, MFCC, MFCC with velocity and acceleration coefficients, and shifted delta cepstrum (SDC) features. In 2013, Bhaskar et al. [48] have carried out LID study using gender independent, gender dependent and hierarchical grouping approaches on 27 Indian languages. Vocal tract features are used to capture the language-specific information. Summary of the prior works related to implicit LID studies mentioned above is provided in Table 2.2.

Table 2.2 Summary of prior works on *implicit* language identification studies

Sl. no.	Features	Models/ Classification techniques	Number of languages and databases	Remarks	Reference
1.	Prosodic and formant features	K-means clustering	Recorded noisy radio signals as database	Formant features are better than the prosody features for LID task	[26]
2.	LPCs and LPCCs	Vector Quantization	20 languages	Accuracy of 65 % is achieved	[9]
3.	Spectral features at syllable level	ANN	5 languages from OGI-MLTS database	Syllabic spectral feature is useful for LID. 95 % accuracy is achieved	[34]
4.	MFCC	GMM	5 languages from OGI-MLTS database	79 % accuracy is achieved	[35]
5.	*Rhythm* at syllable level	GMM	7 languages from MULTEXT corpus	88 % accuracy is achieved	[36]
6.	Production knowledge of vowels and consonants	HMM	5 languages from South Indian Language (SInL) corpus	65 % accuracy is achieved	[38]
7.	Prosody information (rhythm, stress, and intonation)	i-vector based classification	NIST LRE 2009	Prosodic features contain language-specific knowledge	[25]
8.	MFCC	VQ	5 languages	Presence of some CV units is crucial for LID	[11]

(continued)

Table 2.2 (continued)

Sl. no.	Features	Models/ Classification techniques	Number of languages and databases	Remarks	Reference
9.	Weighted linear prediction cepstral coefficients (WLPCC)	AANN	4 languages	93.75 % accuracy is achieved	[44]
10.	MFCC with delta and delta-delta and shifted delta spectrum (SDC) features	Hierarchial based LID system using GMM, HMM and ANN	9 languages	80.56 % accuracy is achieved	[47]
11.	MFCC using block processing, pitch synchronous and glottal closure based approaches	GMM	27 languages from IITKGP-MLILSC database	Glottal closure based approach performs better than other methods	[45]
12.	Prosodic features extracted from syllable, word and phrase levels	GMM	27 languages from IITKGP-MLILSC database	Word level features provide better LID accuracy	[46]

2.3 Prior Works on Excitation Source Features

The LP residual signal has been processed for several speech related tasks such as, speech enhancement, speaker recognition, audio clip classification and emotion recognition. Few works related to the excitation source features are described as below. B. Yegnanarayana and T. K. Raja [49] have analyzed the LP residual signal while the speech signal has been corrupted with additive white noise. It has been observed that, the features obtained from LP residual signal perform well even though the signal to noise ratio (SNR) is low. The excitation source information has also been exploited for robust speaker recognition task. In B. Yegnanarayana et al. [50], have developed a text-dependent speaker verification system using source, supra-segmental and spectral features. The supra-segmental features such as, pitch and duration are explored. Excitation source features extracted from LP residual signal is modeled by auto associative neural network (AANN). Although the supra-segmental and source features individually does not provide good performance. However, combining the evidences from these features improve the performance of the speaker verification system significantly. In this study, Neural network models are used to combinethe evidences from multiple sources of information. In [51], AANN

is proposed for capturing speaker-specific source information present in LP residual signal. Speaker models are built for each vowel to study the speaker information present in each vowel. Using this knowledge an online speaker verification system has been developed. This study shows that, excitation source features also contain significant speaker-specific information. In [52], LP residual signal, its magnitude and phase components are implicitly processed at sub-segmental, segmental and supra-segmental levels to capture speaker-specific information. The speaker identification and verification studies performed using NIST-99 and NIST-03 databases. This study demonstrates that, the segmental level features provide best performance followed by sub-segmental features. The supra-segmental features provide least performance. In [53], segmental level excitation source features are used for language independent speaker recognition study. In [54], LP residual signal has been explored for capturing the audio-specific information. Autoassociative neural network models have been used to capture the audio-specific information extracted from LP residual signal. In [55], the excitation source component of speech has been explored for characterizing and recognizing the emotions from speech signal. In this work, excitation source information is extracted from both LP residual and glottal volume velocity (GVV) signals. In this study, sequence of LP residual samples and their phase information, parameters of epochs and their dynamics at syllable and utterance levels have been used for characterizing emotions. Further, samples of GVV signal and its parameters also explored for emotion recognition task. In [56], a method has been proposed for duration modification using glottal closure instants (GCIs) and vowel onset points (VOPs). The VOPs are computed using the Hilbert envelope of LP residual signal. Manipulation of duration is achieved by modifying the duration of the LP residual with the help of instants of significant excitation as pitch markers. The modified residual is used to excite the time-varying filter. Perceptual quality of the synthesized speech is found to be natural. In [57], GCIs are computed from LP residual signal by using the property of average group-delay of minimum phase signals. The modification of pitch and duration was achieved by manipulating the LP residual with the help of the knowledge of the instants of significant excitation. The modified residual signal was used as excitation signal to the vocal tract resonator. The proposed method is evaluated using waveforms, spectrograms, and listening tests and it is found that, the perceptual quality of synthesized speech has been improved and there were no significant distortion. In K.S. Rao et al. [58], have proposed a time-effective method for determining the instants of significant excitation (GCIs) in speech signals. The proposed methods consist of two phases: (i) at first phase approximate epoch locations using the Hilbert envelope of LP residual signal and (ii) at second phase, accurate locations of the instants of significant excitation is determined by computing the group delay around the approximate epoch locations derived from the first phase. In [59], pitch contours are modified by using the significant instant of excitation and this technique can be used in voice conversion, expressive speech synthesis applications. In [60], excitation source features have been used for voice conversion tasks. The basic goal of the voice conversion

system is to modify the speaker-specific characteristics, keeping the message and the environmental information contained in the speech signal intact. In [60], a neural network models for developing mapping functions at each level has been proposed. The features used for developing the mapping functions are extracted using pitch synchronous analysis. In this work, the instants of significant excitation are used as pitch markers to perform the pitch synchronous analysis. Instants of significant excitation are computed from LP residual signal by using the property of average group-delay of minimum phase signals. In [61], a method has been proposed which is capable of jointly converting prosodic features, spectral envelope and excitation signal maintaining the correlation between them and this method has been used in voice conversion application.

2.4 Motivation for the Present Work

From the prior works related to LID studies mentioned in Sects. 2.1 and 2.2, it is observed that the existing LID systems are mostly developed using spectral features representing the vocal tract system characteristics and prosodic features representing the supra-segmental characteristics of the languages. The excitation source component of speech has still not been explored for LID task. From the literature, it has been observed that excitation source information represented by LP residual signal has been explored for several speech tasks. But, it has not been investigated for language discrimination task. Therefore, in this book, we want to explore excitation source features for language discrimination task. The human speech production system consists of time varying vocal tract resonator and the source for provoking the resonator. Speech sounds are produced as a consequence of acoustical excitation of the human vocal tract resonator. During the production of voiced sounds, the vocal tract is excited by a series of nearly periodic air pulses generated by the vocal cords vibration. State-of-the-art LID systems mostly approximate the dynamics of vocal tract shape and use this vocal tract information for discriminating the languages. However, the demeanor of the vocal folds vibration also changes from one sound unit to another. Although there is a significant overlap in the set of sound units in different languages, but the same sound unit may differ across different languages due to the co-articulation effects and dialects. Hence, we conjecture that, the characteristics of excitation source may contain some language-specific information. In present work, we have explored the excitation source features for capturing language-specific phonotactic information. A theoretical study has been carried out in Sect. 2.4 to support our hypothesis.

Correlation Among the Languages from Excitation Source Point of View

In this section, the significance of the excitation source information for language identification task is shown by their respective correlation coefficients for within and between languages. Correlation determines the degree of similarity between two

Table 2.3 Correlation coefficients across the languages derived from excitation source features

Languages	Correlation coefficients																										
Arunachali	**1.8**	0.83	0.52	0.8	1.02	0.97	0.7	0.49	0.38	0.87	0.93	0.5	0.66	0.42	0.98	0.32	0.78	1.21	0.48	1.21	1.98	0.63	1.1	0.64	0.5	1.4	1.59
Assamese	0.83	**3.26**	0.59	1.18	1.32	0.69	1.68	0.94	0.69	1.09	1.32	0.58	0.91	1.06	1.7	0.78	1.07	1.46	0.41	1.31	1.3	1.4	1.72	1.23	0.83	1.53	1.88
Bengali	0.52	0.59	**2.09**	0.88	0.61	0.57	0.56	0.82	0.82	0.71	0.57	0.63	0.66	0.51	0.54	0.39	0.36	0.74	0.54	0.67	0.87	0.67	0.57	0.36	0.74	0.8	0.68
Bhojpuri	0.8	1.18	0.88	**2.8**	0.72	0.72	0.67	0.95	0.82	1.2	1.26	0.91	1.07	0.79	1.29	0.59	1.19	1.07	0.58	1.21	1.53	1.16	1.53	0.66	1.17	1.39	0.92
Chhattisgarhi	1.02	1.32	0.61	0.72	**3.42**	1.21	1.2	0.74	0.63	0.88	1.53	0.63	0.91	0.88	1.47	0.73	0.91	2.73	0.72	2.18	2.71	0.93	2.23	0.84	0.92	2.74	2.46
Dogri	0.97	0.69	0.57	0.72	1.21	**2.06**	0.62	0.38	0.54	0.71	1.11	0.67	0.69	0.57	1.2	0.53	0.71	1.57	0.59	1.19	1.66	0.55	1.47	0.72	0.64	1.53	1.91
Gojri	0.7	1.68	0.56	0.67	1.2	0.62	**4**	0.76	1.11	0.87	1.37	0.49	0.76	1.19	1.41	1.38	1.04	2.8	0.65	0.91	0.04	0.77	0.72	0.94	0.5	1.63	1.29
Gujarati	0.49	0.94	0.82	0.95	0.74	0.38	0.76	**1.91**	0.64	0.69	0.8	0.55	0.57	0.72	0.75	0.42	0.55	0.87	0.54	0.78	1.63	0.86	0.94	0.64	0.72	1.07	0.92
Hindi	0.38	0.69	0.82	0.82	0.63	0.54	1.11	0.64	**1.67**	0.69	1.12	0.46	0.91	0.93	0.59	0.54	0.47	0.83	0.57	0.75	0.86	0.8	0.93	0.38	0.95	1	0.79
Indian English	0.87	1.09	0.71	1.2	0.88	0.71	0.87	0.69	0.69	**2.53**	0.98	0.55	1.09	0.9	0.75	0.74	0.73	1.28	0.5	1.21	0.87	0.83	1.15	0.93	0.74	0.92	1.05
Kannada	0.93	1.32	0.57	1.26	1.53	1.11	1.37	0.8	1.12	0.98	**2.41**	0.53	1.12	1.05	1.18	0.91	1.02	1.96	0.56	1.13	1.14	0.7	1.63	0.83	0.59	1.62	2.33
Kashmiri	0.5	0.58	0.63	0.91	0.63	0.67	0.49	0.55	0.46	0.55	0.53	**1.28**	0.69	0.46	0.85	0.46	0.65	1.23	0.47	0.92	1.2	0.73	0.71	0.44	0.53	1.01	1.02
Konkani	0.66	0.91	0.66	1.07	0.91	0.69	0.76	0.57	0.91	1.09	1.12	0.69	**3.1**	1.15	1.08	0.57	0.6	0.48	0.8	0.76	2.2	0.67	1.32	0.56	0.94	1.35	0.93
Malayalam	0.42	1.06	0.51	0.79	0.88	0.57	1.19	0.72	0.93	0.9	1.05	0.46	1.15	**2.02**	0.96	1.02	0.6	0.76	0.56	0.84	1.09	1.05	1.5	0.67	0.97	1.13	1
Manipuri	0.98	1.7	0.54	1.29	1.47	1.2	1.41	0.75	0.59	0.75	1.18	0.85	1.08	0.96	**2.9**	0.56	1.48	1.84	0.42	1.54	1.59	1.02	2.25	1.06	1.23	1.69	2.42
Marathi	0.32	0.78	0.39	0.59	0.73	0.53	1.38	0.42	0.54	0.74	0.91	0.46	0.57	1.02	0.56	**1.34**	0.52	1.59	0.41	0.61	0.15	0.54	0.36	0.55	0.5	1.06	0.9
Mizo	0.78	1.07	0.36	1.19	0.91	0.71	1.04	0.55	0.47	0.73	1.02	0.65	0.6	0.6	1.48	0.52	**1.67**	1.35	0.48	1.21	1.24	0.63	1.1	0.51	0.68	1.24	1.59
Nagamese	1.21	1.46	0.74	1.07	2.73	1.57	2.8	0.87	0.83	1.28	1.96	1.23	0.48	0.76	1.84	1.59	1.35	**9.89**	0.41	3.02	0.14	0.99	1.12	1.22	1.09	3.12	3.62
Nepali	0.48	0.41	0.54	0.58	0.72	0.59	0.65	0.54	0.57	0.5	0.56	0.47	0.8	0.56	0.42	0.41	0.48	0.41	**1.27**	0.48	0.88	0.6	0.74	0.43	0.61	1.23	0.53

(continued)

Table 2.3 (continued)

Languages	Correlation coefficients																										
Oriya	1.21	1.31	0.67	1.21	2.18	1.19	0.91	0.78	0.75	1.21	1.13	0.92	0.76	0.84	1.54	0.61	1.21	3.02	0.48	**3.98**	1.84	0.98	1.97	0.92	1.22	2.29	2.23
Punjabi	1.98	1.3	0.87	1.53	2.71	1.66	0.04	1.63	0.86	0.87	1.14	1.2	2.2	1.09	1.59	0.15	1.24	0.14	0.88	1.84	**13.79**	1.79	4.49	0.84	1.57	3.45	3.03
Rajasthani	0.63	1.4	0.67	1.16	0.93	0.55	0.77	0.86	0.8	0.83	0.7	0.73	0.67	1.05	1.02	0.54	0.63	0.99	0.6	0.98	1.79	**1.98**	1.44	0.66	0.62	1.23	1.17
Sanskrit	1.1	1.72	0.57	1.53	2.23	1.47	0.72	0.94	0.93	1.15	1.63	0.71	1.32	1.5	2.25	0.36	1.1	1.12	0.74	1.97	4.49	1.44	**4.82**	0.88	1.18	3.07	2.62
Sindhi	0.64	1.23	0.36	0.66	0.84	0.72	0.94	0.64	0.38	0.93	0.83	0.44	0.56	0.67	1.06	0.55	0.51	1.22	0.43	0.92	0.84	0.66	0.88	**1.67**	0.4	1.19	1.44
Tamil	0.5	0.83	0.74	1.17	0.92	0.64	0.5	0.72	0.95	0.74	0.59	0.53	0.94	0.97	1.23	0.5	0.68	1.09	0.61	1.22	1.57	0.62	1.18	0.4	**2.17**	1.28	0.92
Telugu	1.4	1.53	0.8	1.39	2.74	1.53	1.63	1.07	1	0.92	1.62	1.01	1.35	1.13	1.69	1.06	1.24	3.12	1.23	2.29	3.45	1.23	3.07	1.19	1.28	**6.01**	3.82
Urdu	1.59	1.88	0.68	0.92	2.46	1.91	1.29	0.92	0.79	1.05	2.33	1.02	0.93	1	2.42	0.9	1.59	3.62	0.53	2.23	3.03	1.17	2.62	1.44	0.92	3.82	**5.26**

signals. Suppose that we have two real signal sequences $x(n)$ and $y(n)$ each of which has finite energy. The *cross-correlation* of $x(n)$ and $y(n)$ is a sequence $r_{xy}(l)$, which is defined as follows:

$$r_{xy}(l) = \sum_{n=1}^{p} x(n)y(n-l),\ l = 0, \pm 1, \pm 2, ... \tag{2.1}$$

where, l is the time shift parameter. The x and y are the two signals being correlated. If the signals are identical, then the correlation coefficient is maximum and if they are orthogonal then the correlation coefficient is minimum. When $x(n) = y(n)$, the procedure is known as *autocorrelation* of $x(n)$. From each language database, one male speaker's data of 5 min duration is considered and the LP residual has been extracted. The LP residual is then decimated by factor 4 to suppress the sub-segmental level information and then the LP residual samples are processed in block size of 20 ms with a shift of 2.5 ms which provides segmental level information. To normalize the speaker variability between the languages, the mean subtraction is imposed to all the feature vectors across all languages. Then the *seg* level feature vectors are modeled with GMM for each language. The average *mean* vectors are considered as the signal for a particular language to compute the correlation coefficients. To portray the significance of *seg* level LP residual feature in language discrimination task these correlation coefficients are used. The correlation coefficients between two signals is a sequence of length $(2l - 1)$. The average of the $(2l - 1)$ correlation coefficient values is considered in our work which is shown in Table 2.3. The values of first row of the Table 2.3 indicates the correlation coefficients of first language with respect to itself and other 26 languages. The correlation coefficient within a language has been computed from two different speech utterances spoken by a speaker. The first element of first row indicates the auto-correlation coefficient of first language calculated from the average *mean* vectors of two utterances within one language. The other 26 values of first row represents the cross-correlation coefficients between the first language and other 26 languages. Lower the cross-correlation coefficient value between two languages indicate more dissimilarity between them. We have taken the average of the 26 cross-correlation coefficients from the 2nd column to 27th column of the 1st row which represents average cross-correlation coefficient of first language with respect to other 26 languages. This average cross-correlation coefficient value (0.85) is less than the auto-correlation coefficient value (1.8) which resides in the 1st column of the 1st row. This explains that the *seg* level LP residual feature has significant language discriminative capability. If we analyze the other rows of the Table 2.3, similar characteristics can be observed. This theoretical discussion elicits the significance of the excitation source features in language identification task which is the motivation of the present work.

2.5 Summary

In this chapter, the existing works related to both the *explicit* and *implicit* LID systems have been described. Prior works based on excitation source features are also discussed. It has been observed that, the excitation source component of speech has not been explored for language discrimination task, which is the motivation of present work. Hence, in this work, excitation source information has been explored to capture language-specific phonotactic information for LID task.

References

1. R. Leonard, G. Doddington, Automatic language identification. Technical Report RADC-TR-74-200 (Air Force Rome Air Development Center, Technical Report) August 1974
2. R. Leonard, Language Recognition Test and Evaluation. Technical Report RADCTR-80-83 (Air Force Rome Air Development Center, Technical Report). March 1980
3. A.S. House, E.P. Neuberg, Toward automatic identification of the languages of an utterance. J. Acoust. Soc. Am. **62**(3), 708–713 (1977)
4. K.P. Li, T.J. Edwards, Statistical models for automatic language identification, in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 884–887, April 1980
5. L.F. Lamel, J.L. Gauvain, Cross lingual experiments with phone recognition. in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 507–510, April 1993
6. L.F. Lamel, J.L. Gauvain, Language identification using phonebased acoustic likelihoods, in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 1, , pp. I/293–I/296, April 1994
7. Y. Muthusamy, R. Cole, M. Gopalakrishnan, A segment-based approach to automatic language identification, in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 1, pp. 353–356, April 1991
8. K.M. Berkling, T. Arai, E. Bernard, Analysis of phoneme based features for language identification, in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. I/289–I/292, April 1994
9. R.C.F. Tucker, M. Carey, E. Parris, Automatic language identification using sub-word models, in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 1, pp. I/301–I/30, April 1994
10. M.A. Zissman, E. Singer, Automatic language identification of telephone speech messages using phoneme recognition and N-gram modeling, in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 1 pp. I/305–I/308, (1994)
11. S. Kadambe, J. Hieronymus, Language identification with phonological and lexical models, in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5, pp. 3507–351, May 1995
12. Y. Yan, E. Barnard, An approach to automatic language identification based on language-dependent phone recognition, in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5, pp. 3511–3514, May 1995
13. J. Navratil, W. Zuhlke, Phonetic-context mapping in language identification. Eur. Speech Commun. Assoc. (EUROSPEECH) **1**, 71–74 (1997)
14. T.J. Hazen, V.W. Zue, Segment-based automatic language identification. J. Acoust. Soc. Am. **101**, 2323–2331 (1997)
15. K. Kirchhoff, S. Parandekar, Multi-stream statistical n-gram modeling with application to automatic language identification, in *European Speech Communication Association (EUROSPEECH)*, pp. 803–806, (2001)

16. T. Gleason, M. Zissman, Composite background models and score standardization for language identification systems, in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 1, pp. 529–532 (2001)
17. V. Ramasubramanian, A.K.V.S. Jayram, T.V. Sreenivas, Language identification using parallel sub-word recognition - an ergodic HMM equivalence, *European Speech Communication Association (EUROSPEECH)* (Geneva, Switzerland), September 2003
18. J. Gauvain, A. Messaoudi, H. Schwenk, Language recognition using phone latices, in *International Speech Communication Association (INTERSPEECH)*, pp. 25–28 (2004)
19. W. Shen, W. Campbell, T. Gleason, D. Reynolds, E. Singer, Experiments with lattice-based PPRLM language identification, in *Speaker and Language Recognition Workshop*, pp. 1–6 (2006)
20. H. Li, B. Ma, C.H. Lee, A vector space modeling approach to spoken language identification. IEEE Trans. Audio Speech Lang. Process. **15**(1), 271–284 (2007)
21. K.C. Sim, H. Li, On acoustic diversification front-end for spoken language identification. IEEE Trans. Audio Speech Lang. Process. **16**(5), 1029–1037 (2008)
22. R. Tong, B. Ma, H. Li, E.S. Chng, A target-oriented phonotactic front-end for spoken language recognition. IEEE Trans. Audio Speech Lang. Process. **17**(7), 1335–1347 (2009)
23. G.R. Botha, E. Barnard, Factors that affect the accuracy of text-based language identification. Comput. Speech Lang. **26**(5), 307–320 (2012)
24. N. Barroso, K. Lopez de Ipina, C. Hernandez, A. Ezeiza, M. Grana, Semantic speech recognition in the Basque context Part II: language identification for under-resourced languages. Int. J. Speech Technol. **15**(1), 41–47 (2012)
25. S.M. Siniscalchi, J. Reed, T. Svendsen, C.-H. Lee, Universal attribute characterization of spoken languages for automatic spoken language recognition. Comput. Speech Lang. **27**(1), 209–227 (2013)
26. J.T. Foil, Language identification using noisy speech, in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 861–864, (1986)
27. F. Goodman, A. Martin, R. Wohlford, Improved automatic language identification in noisy speech, in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 1, pp. 528–531, May 1989
28. M. Sugiyama, Automatic language recognition using acoustic features, in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 813–816, May 1991
29. D. Morgan, L. Riek, W. Mistretta, C. Scofield, P. Grouin, F. Hull, Experiments in language identification with neural networks. Int. Joint Conf. Neural Netw. **2**, 320–325 (1992)
30. M. Zissman, Automatic language identification using gaussian mixture and hidden markov models, in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2, pp. 399–402, April 1993
31. D.A. Reynolds, R.C. Rose, Robust text -independent speaker identification using gaussian mixture speaker models. IEEE Trans. Audio Speech Lang. Process. **3**(1), 72–83 (1995)
32. S. Itahashi, J. Zhou, K. Tanaka, Spoken language discrimination using speech fundamental frequency, in *International Conference on Spoken Language Processing (ICSLP)*, pp. 1899–1902, (1994)
33. I. Shuichi, D. Liang, Language identification based on speech fundamental frequency, in *European Speech Communication Association (EUROSPEECH)*, pp. 1359–1362 (1995)
34. K.P. Li, Automatic language identification using syllabic spectral features, in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 1, pp. I/297–I/300, April 1994
35. F. Pellegrino, R. Andre-Obrecht, An unsupervised approach to language identification, in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2, pp. 833–836, Mar 1999
36. J.L. Rouas, J. Farinas, F. Pellegrino, R. Andr-Obrecht, Rhythmic unit extraction and modelling for automatic language identification. Speech Commun. **47**, 436–456 (2005)
37. J.L. Rouas, Automatic prosodic variations modeling for language and dialect discrimination. IEEE Trans. Audio Speech Lang. Process. **15**(6), 1904–1911 (2007)

38. A. Sangwan, M. Mehrabani, J. Hansen, Automatic language analysis and identification based on speech production knowledge, in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 5006–5009, March 2010
39. D. Martinez, L. Burget, L. Ferrer, N. Scheffer, i-vector based prosodic system for language identification, in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4861–4864, March 2012
40. J. Balleda, H.A. Murthy, T. Nagarajan, Language Identification from Short Segments of Speech, in *International Conference on Spoken Language Processing (ICSLP)*, pp. 1033–1036, October 2000
41. T. Nagarajan, Implicit system for spoken language identification, Ph.D. dissertation, Indian Institute of Technology Madras, India (2004)
42. A.K.V.S. Jayaram, V. Ramasubramanian, T.V. Sreenivas, Language identification using parallel sub-word recognition, in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 32–35, April 2003
43. L. Mary, B. Yegnanarayana, Autoassociative neural network models for language identification. in *Internatioanl Conference on Intelligent Sensing and Information Processing*, pp. 317–320 (2004)
44. L. Mary, Multilevel implicit features for language and speaker recognition, Ph.D. dissertation, Indian Institute of Technology Madras, India (2006)
45. K.S. Rao, S. Maity, V.R. Reddy, Pitch synchronous and glottal closure based speech analysis for language recognition. Int. J. Speech Technol. (Springer) **16**(4), 413–430 (2013)
46. V.R. Reddy, S. Maity, K.S. Rao, Identification of indian languages using multi-level spectral and prosodic features. Int. J. Speech Technol. (Springer) **16**(4), 489–511 (2013)
47. S. Jothilakshmi, V. Ramalingam, S. Palanivel, A hierarchical language identification system for Indian languages. Digital Signal Process. (Elsevier) **22**(3), 544–553 (2012)
48. B. Bhaskar, D. Nandi, K.S. Rao, Analysis of language identification performance based on gender and hierarchial grouping approaches, in *International Conference on Natural Language Processing*, December 2013
49. B. Yegnanarayana, T.K. Raja, Perfoemance of linear prediction analysis on speech with additive noise, in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (1977)
50. B. Yegnanarayana, S.R.M. Prasanna, J. Zachariah, C. Gupta, Combining evidence from source, suprasegmental and spectral features for a fixed-text speaker verification system. IEEE Trans. Audio Speech Lang. Process. **13**(4), 575–582 (2005)
51. C.S. Gupta, S.R.M. Prasanna, B. Yegnanarayana, Autoassociative neural network models for online speaker verification using source features from vowels, in *IEEE International Joint Conference Neural Networks*, May 2002
52. D. Pati, S.R.M. Prasanna, Subsegmental, segmental and suprasegmental processing of linear prediction residual for speaker information. Int. J. Speech Technol. (Springer) **14**(1), 49–63 (2011)
53. D. Pati, D. Nandi, K. Sreenivasa Rao, Robustness of excitation source information for language independent speaker recognition, in *16th International Oriental COCOSDA Conference*, Gurgoan, November 2013
54. A. Bajpai, B. Yegnanarayana, Exploring features for audio clip classification using LP residual and AANN models, in *International Conference on Intelligent Sensing and Information Processing*, pp. 305–310, January 2004
55. K.S. Rao, S.G. Koolagudi, Characterization and recognition of emotions from speech using excitation source information. Int. J. Speech Technol. (Springer) **16**, 181–201 (2013)
56. K.S. Rao, B. Yegnanarayana, Duration modification using glottal closure instants and vowel onset points. Speech Commun. **51**(12), 1263–1269 (2009)
57. K.S. Rao, B. Yegnanarayana, Prosody modification using instants of significant excitation. IEEE Trans. Audio Speech Lang. Process. **14**(3), 972–980 (2006)
58. K.S. Rao, S.R.M. Prasanna, B. Yegnanarayana, Determination of instants of significant excitation in speech using Hilbert envelope and group delay function. IEEE Signal Process. Lett. **14**(10), 762–765 (2007)

59. K.S. Rao, Unconstrained pitch contour modification using instants of significant excitation. Circuits Syst. Signal Process. (Springer) **31**(6), 2133–2152 (2012)
60. K.S. Rao, Voice conversion by mapping the speaker-specific features using pitch synchronous approach. Comput. Speech Lang. **24**(3), 474–494 (2010)
61. R. Hussain Laskar, K. Banerjee, F. Ahmed Talukdar, K. Sreenivasa Rao, A pitch synchronous approach to design voice conversion system using source-filter correlation. Int. J. Speech Technol. (Springer) **15**(3), 419–431 (2012)

http://www.springer.com/978-3-319-17724-3

Language Identification Using Excitation Source Features

Rao, K.S.; Nandi, D.

2015, XII, 119 p. 19 illus., 3 illus. in color., Softcover
ISBN: 978-3-319-17724-3