Comparative Study of Techniques for Spoken Language Dialect Identification

Ashwini G. Pawar
School of Computer Sciences,
North Maharashtra University,
Jalgaon, India

Nita V. Patil
School of Computer Sciences,
North Maharashtra University,
Jalgaon, India

ABSTRACT
Identifying variations in spoken language resulting from regional or socioeconomic factors, known as dialect identification, is a significant problem in natural language processing and linguistics. This study contrasts various dialect identification approaches and evaluates their effectiveness. Techniques include deep learning, transfer learning, classical acoustic feature analysis, machine learning (ML) algorithms, phonetic and phonological analysis, lexical and grammatical feature extraction, prosodic analysis, and phonological analysis. Meticulous application of these techniques is applied to a hand-picked dataset of multiple dialects, with their performance assessed using accepted evaluation measures. These findings reveal that different techniques capture dialectal variations differently. Phonetic and phonological analysis excels at detecting minute pronunciation changes, while acoustic feature-based ML demonstrates resilience in dialect discrimination. Lexical and grammatical factors effectively recognize small differences in vocabulary and grammar usage. Prosodic features enhance dialect identification through intonation and rhythm patterns. Moreover, deep learning models showcase their capacity to learn intricate patterns from large datasets, and transfer learning techniques are effective in scenarios with limited dialect-specific data. Multilingual and cross-lingual approaches leverage shared linguistic properties for enhanced identification accuracy. Ensemble methods harness the strengths of multiple techniques, resulting in improved overall performance. This study underscores the significance of a diversified approach to dialect identification, with the choice of technique depending on factors such as available resources, data availability, and dialect complexity.

General Terms
Natural Language Processing, Language Dialect Identification, Speech Processing.

Keywords
Dialect identification, deep learning, acoustic analysis, phonological analysis, transfer learning, multilingual approaches, ensemble methods.

1. INTRODUCTION
Speech signal processing and dialect identification are exciting fields that combine linguistics and signal processing to understand and distinguish between regional or social variations in spoken language. These two aspects are as follows.

Speech Signal Processing- Using diverse mathematical and computational techniques, speech signal processing entails analyzing, modifying, and interpreting voice signals. The objective is to extract relevant information from the speech.

Linguistic Aspects of Speech and Dialect Identification- Dialect identification is fundamentally linked to linguistic variation. Additionally, dialects are identified as geographical and social variants of a language by the speaker's syntactic, lexical, and phonological shift in pronunciation (accent) [1]. To better understand the diversity and evolution of languages, linguists analyze these variances.

Other than Spoken languages, people use language dialects for communication, and Human beings can distinguish between spoken languages and their dialects as a part of human intelligence [2]. People frequently say things like "it sounds like Konkani Marathi" when describing new languages with languages they are familiar with. Even though these conclusions are less precise when making selections for identification, they nonetheless demonstrate how humans use lingual information at various levels to distinguish between diverse groups of spoken languages and their varied dialects.

A dialect is a variety of speech associated with a place with a literary language or speech pattern distinct from the culture in which it is used. For example, Konkani Marathi (spoken by people of Konkan) and Vidarbhi Marathi (spoken by people of Vidarbh) are dialects of Marathi. Language Dialect Identification automatically recognizes the dialect from spoken utterances and a speaker's regional dialect within a predefined language. It is associated with features like acoustic, spectral, and suprasegmental (Prosodic). Separating these elements from spoken signals can readily determine the language's dialect. Moreover, the spoken language sample's phonotactic, spectral, and prosodic characteristics can provide enough about the speaker's native tongue.

Combining speech signal processing with linguistic knowledge allows for a comprehensive and accurate approach to dialect identification. It enables us to analyze the acoustic properties of speech, which is discussed in a further section of the paper, while considering the underlying linguistic variations that give rise to different dialects.

This paper reviews Spoken Language Dialects for various Indian and non-Indian languages and the techniques and stages for language dialect identification through speech signal processing in section 2. This paper is organized into six sections. Section 3 discusses speech signal processing and analysis. Section 4 explains the linguistic features of Language Dialects and focuses on feature extraction techniques. Section 5 elaborates on various applications of understanding speech signal processing from a language aspect, including spoken language dialect. In Section 6, the paper is discussed, and in Section 7, the paper concludes.

2. BRIEF SURVEY OF LITERATURE ON SPOKEN LANGUAGE DIALECT IDENTIFICATION
Jooyoung Lee et al. [3] developed Korean dialect identification

(K-DID), a computational model for K-DID trained on a large-scale corpus, learning intonation patterns based on theories and previous works from Korean dialectology. The study aims to make a classifier to detect dialectal patterns among non-dialectal intervals.

Sunil Patil et al.[4] employed the Hidden Markov Model Toolkit (HTK) to achieve speaker-independent speech recognition for the Marathi dialect varhadi through isolated word recognition. M. Nanmalar et al.[5] proposed categorizing literary and colloquial Tamil dialects using acoustical properties, reducing the reliance on language-specific linguistic tools. Their method, which utilizes Mel Frequency Cepstral Coefficient (MFCC) features and Gaussian Mixture Models (GMM) for classification, stands out for its adaptability to diverse languages without annotated corpora. Using a neural network-based Q-learning algorithm, Himanish Shekhar Das and Pinki Roy [6] devised a way to get the best prosodic feature and finish the classification in parametric excitation source data to identify Indian languages. Prosodic features are swiftly extracted from speech signals using the suggested method. The deep neural network-based Q-learning (DNNQL) technique was utilized to determine a language's class label, and the Fourier transform was applied to handle non-stationary data. A database of Indian language recordings proves the strategy works.

L. Joyprakash and Tanvira Ismail [7] Singh created the Speech corpus necessary for identifying the two varieties of Assamese, an Indian language, and the method for doing it. Dialects and languages have been detected using the Gaussian Mixture Model and its Universal Background Model.

Using the acoustic characteristics of speech, Nagaratna B. Chittaragi and Shashidhar G. Koolagudi [8] devised a word-based dialect categorization system. They used word-level spectral and prosodic acoustic variables to apply this method to the Intonational Variations in English (IViE) corpus containing nine British English dialects. They used SVM and tree-based extreme gradient boosting (XGB) ensemble approaches to find language categorization patterns.

Automatic Language Identification for Seven Indian Languages [9] was made based on a plan by Chithra Madhu et al. It makes use of prosodic information and language-dependent phonotactic elements. The front end of the phonotactic-based Language Identification (LID) system, which turns spoken words into a series of phonetic symbols, was a "phonetic engine." When syllable boundaries are found, the phones that make up that syllable are grouped, and phonotactic rules are used to get syllables. The phonotactic feature vectors are a numerical representation of two successive syllables. Prosodic feature vectors are formed by connecting three consecutive syllables. The language detection procedure uses a multilayer feed-forward neural network (NN) classifier. The ANN classifier was trained with two hours of data from each of the seven languages. The study examines Assamese, Bengali, Hindi, Telugu, Urdu, Punjabi, and Manipuri. Aman Ankit et al. investigated the usage of the Sphinx engine for automatic voice recognition [10]. Speech recognition research-based library called Sphinx. In order to improve Automatic Speech Recognition (ASR), Ashok Shigli et al. [11] concentrate on automatically identifying the dialect or accent of a speech by a speaker of South Indian English. They use the Mel Frequency Cepstral Coefficient (MFCC) algorithm.

Reena Chaudhari et al. [12] have studied how Hindi has influenced languages like Marathi, Marwadi, and Urdu. These models use MFCC and LPC with acoustic Features to recognize accents in the Hindi language, along with a few additional features. In research by Gang Liu and John H. L. Hansen, the terms "accent" and "dialect" were defined as the patterns of pronunciation and vocabulary of a language used by a population of native speakers in a specific geographical region [13]. Raymond W. M. Ng et al. [14] proposed a feature selection strategy to enhance the performance of the Prosodic Language Identification system. They applied this strategy to evaluate and select prosody-related factors for language identification (LID) tasks, introducing novel frontend and backend designs. The front end included Parallel Phone Recognition (PPR) and Universal Phone Recognition (UPR) to convert speech into segment units. In the backend, they utilized ASM-derived feature vectors based on ASM data and their co-occurrences, implementing a novel Vector Space Modeling (VSM) technique [15] for distinguishing spoken languages.

The following summary table explains techniques for Indian and Non-Indian language Dialect Identification.

Table 1. Summary of techniques for Indian and Non-Indian languages Dialect Identification

Languages	Authors	Year	Corpus	Techniques	Nature of work
Korean Language	Jooyoung Lee, KyungWha Kim, Minhwa Chung [1]	2021	A large-scale audio corpus comprising over 2,500 Korean speakers, the Korean Standard Speech Database (KSS-DB), contains 500 hours of audio.	Baseline for Dialect Identification in Korea. Bidirectional LSTM Network with attention-based. One versus Rest (OvR) and One versus One (OvO) evaluation methods	Identify Korean Dialects
Gujarati	Deepang Raval, Vyom Pathak, Muktan Patel, Brijesh Bhatt [16]	2020	3,075 test examples in Gujarati	A BERT-based post-processing method using a 4-gram word-level language model (WLM) and a bi-gram character-level language model (CLM)	By examining the distribution and occurrence of the single-letter mistake words, it was possible to evaluate the effectiveness of the decoding method and post-processing.
Kannada Dialects	Nagaratna B. Chittaragi, Shashidha G. Koolagudi1[17]	2019	Vowel dataset of Native speakers of Kannada	Acoustic Phonetic features	Acoustic metrics, including formant frequencies (F1-F3),1, and prosodic qualities [energy, pitch

					(F0), and duration], are used to characterize dialectal cues.
German Language	Alina Maria Ciobanu, Shervin Malmasi, Liviu P. Dinu[18]	2018	Speakers from Basel, Bern, Lucerne, and Zurich were included in the dataset.	Classifier Ensembles	Identification of German dialects.
Arabic Dialects	Shervin Malmasi, Marcos Zampieri[19]	2017	Arabic Dialects in South Africa	ASR Transcripts and i-vector	Using transcriptions of the audio and text, determine the dialect of Arabic utterances.
Assamese Language	T. Ismail, L. Joyprakash Singh [7]	2017	Newly created corpus by researchers	Models for Gaussian mixtures and those with a universal background	By employing spectral features, an Assamese dialect can be identified.
British English	Nagaratna B. Chittaragi, Shashidhar G. Koolagudi [8]	2017	Nine collections representing nine different British Isles regions.	SVM and Ensemble Methods	Classification of Word Level Dialects Based on Acoustic Features.
Bengali, Hindi, Telugu, Urdu, Assamese, Punjabi and Manipuri	Chithra Madhu, Anu George and Leena Mary [9]	2017	Speech Utterance of mentioned languages	Neural network (NN) classifier	Development of language Identification system.
South Indian English accent recognition by Telugu, Kannada, Tamil, and Marathi speakers in south region	Ashok Shigli, Ibrahim Patel and K Srinivasa Rao[11]	2016	Corpus of South Indian English spoken by Telgu, Marathi, Kannada and Tamil Speakers	Hidden Markov Model (HMM) with Mel frequencyCepstral Coefficients (MFCC)	English spoken by South Indians is automatically recognized for accent and dialect.
Hindi	Shweta Sinha, Aruna Jain and S. S. Agrawal [20]	2015	Dataset of Hindi, Urdu, Marwadi, and Marathi	MFCC and LPC with acoustic features	Determine how Hindi has influenced other Indian languages, such as Marathi, Marwadi, and Urdu, with accents and other characteristics.
Hindi Dialects	Shweta Sinha, Aruna Jain and S. S. Agrawal [21]	2015	Corpus of Hindi dialects by Native speakers	Perceptual Linear Prediction Coefficients (PLP) and MFCC	Develop a Hindi dialect identification system based on acoustic-phonetic features.
Marathi	Aman Ankit, Sonu Kumar Mishra [22]	Sep. 2016	Newly created corpus by Researchers	SPHINX(Speech Recognizer Tool)	Establish the Marathi language corpus and investigate the Sphinx engine's potential for ASR.

3. SPEECH SIGNAL

Dialect identification and speaker recognition are two of the most commonly used applications in speech processing, which is used to identify any language or its dialect. The first stage is to process and analyze the voice signal because it initially takes the shape of an image. Speech signals have some essential traits that are mainly employed for segmenting them. These traits are grouped into two categories, and they are as follows.

3.1.Temporal –Time domain (Based) signal features)

Calculating speech segment extraction uses these criteria. Short-time energy and zero crossing are the most significant components of speech signal analysis for predicting what will happen. Extraction and scientific interpretation of the signal's energy, zero crossing rate, maximum amplitude, and minimum energy are simple [23].

- Short-Time Energy - Short-term energy is essential to speech. Signal strength is energy. Signals naturally change energy
- Short-Time Analysis- The classification of speech signal segments into voiced (with vocal chord vibrations), unvoiced (without vibrations), silent (absence of speech output), and noise (interference) determines the duration of an audio frame. Identifying these segments using energy-based analysis can be difficult, especially for unvoiced speech, mainly because the speech signal zero crossing rate is low.
- Short-Time Zero Crossing- This details the quantity of zero crossings found in a specific signal. If there are more zero crossings in a signal, the signal changes quickly and may thus contain high-frequency information. The highest correlation coefficient and plosion index are two examples of time domain (temporal) features that are comparatively more noise-resistant.2016 (Poornima)

[24].

3.2.Spectral-frequency domain (Based) Signal Features

The speaker-specific vocal tract information includes features such as the spectral centroid, spectral flux, band energies, formants, spectrum, and cepstral coefficients. The Fourier Transform permits the extraction of the fundamental frequency, frequency components, spectral density, and spectral roll-off, essential for determining rhythm, pitch, and other speech features from a time-based signal translated into a frequency-based representation.

Beginners most frequently employ the direction of frequency modulation, such as raising or lowering the level, when studying spoken words as a spectral feature. This supports the study of intonation and tone. Pitch, stress, power spectral density, vowel duration, rhythm, and intonation patterns are further characteristics of speech signals. These features are primarily used in segmentation, which is recognized and significant to identify language dialects [25].

3.3.The Dialect Recognition system involves the following stages

Different types of information that reveal a speaker's identity and dialects are contained in spoken input data. This contains information about the speaker's unique vocal tract, excitation source, and behavioral characteristics, which aids in identifying the spoken language dialects of the speaker (Ma & Li, 2006) [26]. Language-specific information is acquired from feature, token, and prosody levels [27], and speaker recognition signals also include behavioral and linguistic characteristics. These cues cover token distribution and transition variations, prosodic properties, and phoneme implementation, allowing language identification algorithms to classify inputs at different linguistic levels [28]. Short-term energy is essential to speech. Signal strength is energy. Signals naturally change energy.

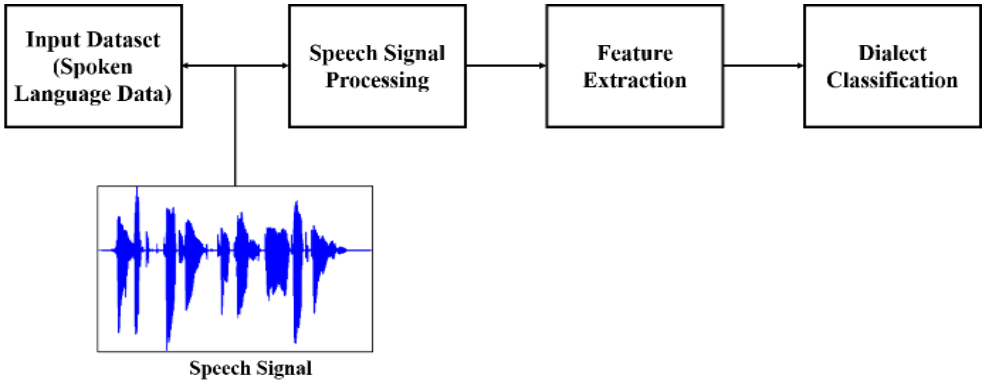

Fig 1: Generalized block diagram of Spoken Dialects Recognition

3.3.1. Input Dataset

The dataset for dialect recognition is a speaker's spoken audio file, which consists of the following types of speech:

- Isolated Word -. It only accepts one word at once.
- Connected Word – Similar to isolated words, but with a slight delay between each of the many words given to the system for running as isolated words.
- Continuous Speech – Without any pause, it enables the user to speak almost naturally.
- Spontaneous Speech - The ability to manage a range of natural qualities, such as words running together, mispronunciation, and erroneous assertions, which are challenging to read, is present in this natural and spontaneous language.

Some of the critical steps involved in speech signal processing for dialect identification include Pre-processing - The raw speech signal is pre-processed to remove background noise, normalize the amplitude, and transform it into an analytically valuable representation, like the Mel- frequency cepstral coefficients (MFCCs)

- Feature Extraction - From the pre-processed voice signal, pertinent features are retrieved. Since they can handle pronunciation changes and capture the spectrum properties of speech, MFCCs are frequently used.
- *Acoustic Modeling* - Based on the retrieved features, acoustic models are created to reflect the speech characteristics of various dialects. For this, ML methods like Deep Neural Networks (DNNs) and Gaussian Mixture Models (GMMs) are frequently used.
- *Dialect Identification*- After being trained, the acoustic models can be used to determine the dialect of a particular speech segment. The system assigns the most likely dialect label by comparing the acoustic characteristics of the input speech with the models.

3.3.2. Speech Signal Processing

Voice processing involves physics, computer science, pattern recognition, and linguistics [29]. Initially, it separates voiced and unvoiced speech portions. Speech pre-processing enhances speech recognition systems, including pre-emphasis, vocal tract length normalization, voice activity recognition, noise removal, framing, and windowing. Reducing noise, a common issue in audio, is crucial for speech processing performance [30]. Noise reduction techniques, such as zero-crossing rate and energy calculations, distinguish voiced and voiceless speech segments, resulting in improved noise reduction [31].

4. FEATURES OF SPEECH

These features collectively form the foundation of spoken language and are crucial for understanding phonetics and the production and interpretation of speech.

4.1 Linguistic Features in Speech Signal

Exploring Linguistic Features in Speech Signal Processing involves analyzing and extracting various linguistic characteristics from speech signals to gain insights into the

underlying linguistic content and improve the performance of speech-processing tasks. Applications like speech recognition, speaker identification, emotion recognition, language modeling, and many others heavily rely on these linguistic properties. Here are some essential linguistic elements that speech signal processing frequently examines.

- Phonetics and Phonemes: Phonetics studies speech sounds and their acoustic characteristics, while phonemes (represented by IPA symbols) are the smallest units influencing word meanings and are crucial for voice recognition systems.
- Prosody: Prosody encompasses speech rhythm, intonation, and stress patterns. It is essential for applications like emotion recognition, voice recognition, and speech synthesis, enabling emotional and sentiment analysis for identifying the speaker's state.
- Pitch and Fundamental Frequency (F0): F0 represents a signal's fundamental frequency, the lowest frequency component, while pitch is the perceived voice frequency. They play key roles in speaker recognition and emotion perception.
- Mel-Frequency Cepstral Coefficients (MFCCs): MFCCs are crucial speech processing features used in Automatic Speech Recognition (ASR) systems, providing speech spectrum information based on the Mel scale.
- Formants: Formants reveal details about vocal tract structure, and changes in formant frequencies during speech sounds are known as formant transitions. These transitions are analyzed to identify transitional boundaries in ASR systems.
- Language Identification: Linguistic features are used for distinguishing between languages in multilingual speech processing systems. Accurate detection of silence and spoken segments is vital for voice activity detection (VAD) and speech segmentation.
- Phonetic Features: Phonetic features refer to distinctive sounds or phonemes in a language, and they help linguists define and distinguish dialects.
- Lexical Features: Lexical features, such as regional vocabulary and word selections, differentiate dialects, as many dialects vary by using particular terms.
- Syntactic Features: Differences in grammar and sentence construction between dialects are recognized by identifying unique syntactic traits common to specific dialects.
- Sociolinguistic Factors: Sociolinguistics explores the interaction between language and social constructs, including how dialects are connected to social groups, geographical areas, or demographics. Incorporating sociolinguistic aspects enhances the accuracy of dialect recognition algorithms.

Studying the acoustical properties of speech is known as acoustic analysis. It involves analyzing waveforms, performing FFT or LPC analyses, measuring formant frequencies, and other physical characteristics of spoken language. These spectral features—MFCC, SDCC, spectral roll-off, flatness, contrast, bandwidth, flux, and centroid provide information about the energy distribution in speech signals and support sound separation and speech analysis.

Rhythm, intonation, and structural components all fall under prosody in spoken language. In order to demonstrate prosodic prominence, various auditory parameters are used, including fundamental frequency, length, energy, and intensity. In perceptual research, syllable rhythm can also distinguish between languages and dialects. Contrarily, prosodic qualities are understood to communicate various information, including lexical tones, speaking styles, etc.

4.2 Feature Extraction

For effective Automatic Speech Recognition (ASR), dialect recognition systems use feature extraction techniques like MFCCs to distinguish between speech patterns, emphasizing traits that are simple to measure, impervious to replication, adaptable to various environments, temporally consistent, and naturally present in speech [32].

This step removes essential parts of speech and extracts relevant information from speech signals to recognize speech and language dialect.

Although it is possible to detect languages or dialects directly from digitized waveforms, feature extraction decreases the dimensionality of voice input, improving signal strength and maintaining discriminative power. Feature extraction improves the input for greater accuracy by lowering unpredictability and dimensionality, while theoretically, digitized waveforms could enable instant speech recognition [6].

In this section, the speech component of the voice signal is taken out and transformed into a digital signal. There are many methods for extracting features from a voice signal.

- MFCC: Widely used in audio processing, Mel Frequency Cepstral Coefficients (MFCC) derive essential features from sound signals via short-term power spectrum analysis.
- MFCC Applications: MFCC is crucial for speech recognition, speaker identification, and dialect identification, enhancing audio signal analysis.
- LPC and Sociolinguistics: Linear Predictive Coding (LPC) is employed in sociolinguistics to explore the relationship between dialects and social characteristics. Adding sociolinguistic factors improves dialect detection accuracy. LPC, combined with cepstral analysis (Linear Prediction Cepstral Coefficients (LPCC), enhances speech and audio feature extraction.
- LPCC: LPCC extracts informative features from audio signals, which is especially beneficial for applications like speaker identification and speech recognition. It bolsters LPC by improving discriminative power and noise robustness through cepstral analysis.
- PCA: Principal Component Analysis (PCA) is commonly used for feature extraction and dimensionality reduction in statistics and data analysis. It generates uncorrelated variables (principal components) from potentially related features, first capturing the most significant data variance and subsequent components capturing decreasing variances.

The following table shows methods, uses, properties, and implementation procedures to extract various features from speech signals.

Table 2. List of Feature extraction techniques along with their properties

Sr. No.	Method	Use	Property	Procedure of Implementation
1	Principal (PCA) Component Analysis	Speech covariance neuroscience analysis	method of nonlinear feature extraction, quick, eigenvector-based linear map	The eigenvector base method, or kahuna-Loeve expansion, is a conventional technique that works well with Gaussian data.
2	Linear (LDA)\ Discriminate Analysis	It identifies the transformation that minimizes the within-class dispersion while maximizing the scatter between classes.	nonlinear feature extraction technique, Fast, eigen vector-based supervised linear map	superior to PCA in terms of categorization
3	Independent Component Analysis (ICA)	It is used to separate multivariate signals into addictive subcomponents.	Linear map, iterative non-Gaussian is nonlinear feature extraction	For de-mixing non-Gaussian distributed sources (features), blind course separation is utilized.
4	Linear Predictive Coding (LPC)	Used in the Shorten, MPEG, FLAC, ALS, and SILK formats, encoding for high-quality speech at a low bit rate.	Method for extracting static features with 10 to 16 lower order coefficients.	It is utilized for lower-order feature extraction.
5	Cepstral Analysis	The speech is divided into its source and system components using it.	Power spectrum, Static feature extraction	utilized to depict the spectral envelope
6	Mel-frequency scale Analysis	Speech Recognition and Audio Retrieval.	Fourier remodel calculations, a method for extracting static features, and a spectrum analysis.	A fixed resolution is used for spectral analysis along the subjective Mel-frequency. Scale.
7	Filter bank approach	It incorporates psychoacoustic findings, which leads to better recognition performance.	Filters tuned required frequencies.	Architectural investigation and logic design are the two critical stages of the implementation process.
8	Mel-Frequency Cepstrum (MFCCs)	This extracted feature from the audio signal used as a base model input performs better than the raw audio signal.	Fourier Analysis is used to compute the power spectrum.	To find technique. features
9	Kernel-based feature extraction method	It is used for pattern analysis. Using a linear classifier, it solve nonlinear problems. It projects data into high-dimensional space.	Alterations that are not linear	Dimensionality reduction eliminates redundant features, improves classification error, and increases classification performance.
10	Wavelet	It provides simultaneous time and frequency localization.	Superior to Fourier Transform for time resolution	It does so by swapping out the Fourier Transform's fixed bandwidth for one proportional to frequency, enabling higher time resolution at high frequencies.
11	Dynamic feature extractions i)LPC ii)MFCCs	LPC's primary objective is the frame-based analysis of the input speech signal to create experimental vectors.	Acceleration and delta coefficients, or the usual MFCC and LPC coefficients' II and III order derivatives.	Runtime features or dynamic features use it.
12	Spectral subtraction	It is used to estimate the noise spectrum during speech pauses.	Robust feature extraction technique.	It is utilized based on spectrograms.
13	Cepstral mean subtraction	It is applied to feature vectors to compensate for the convolute effects of the transmission channel.	Extraction of robust features	It is identical to MFCC anduses the Mean Statically Parameter.
14	RASTA filtering	It changes to use mean subtractionby implementing a moving average filter.	To Quiet Noisy Speech	A feature is being discoveredin noisy data.
15	(Compound Method) IntegratedPhoneme subspace method	In order to represent phoneme information effectively, it gathers correlation data from different phoneme subspaces and reconstructs feature space.	PCA+LDA+ICA-based transformation	greater accuracy

The above table details various nonlinear, static, and dynamic feature extraction techniques used in voice recognition. The PCA, LDA, and ICA combination is exceptionally accurate compared to RASTA filtering for handling loud speech. Different tactics have been developed over time to improve the accuracy of voice recognition systems. The table presents various speech recognition methods, tackling computational issues and assisting in identifying spoken language dialects, such as sub-band-based, dynamic temporal warping, fuzzy logic, wavelet-based, and optimization-based approaches. Spectrogram, phonetic, and linguistic data are used in an inventive Artificial Intelligence approach to pattern recognition

and acoustic-phonetic techniques, greatly enhancing recognition accuracy [33].

A summary of several techniques for speech and audio signal processing is given in the above table, including rapid feature extraction using PCA, categorization using LDA, source separation using ICA, speech encoding using LPC, and other spectral analysis methods. These techniques are used to improve speech processing applications by performing tasks like feature extraction, noise reduction, and phoneme representation. The following table lists numerous voice recognition techniques, each with unique benefits and downsides. Acoustic Phonetic Recognition accelerates word processing but is slow for commercial use. Pattern recognition simplifies word-to-word matching but faces template storage and adaptation challenges. Template-based methods reduce errors for single words but are resource-intensive. Dynamic Time Warping offers flexibility but demands precise constraints and complex computations. Vector quantization reduces data but lacks granularity. Statistical approaches like Hidden Markov Models handle large vocabularies but require significant resources. Artificial Neural Networks excel at computations but struggle with larger vocabularies, high costs, and extended training. Wavelet-based methods offer localization but lack phase and directionality. Fuzzy Logic is robust but requires precise setup and testing. Optimization algorithm-based solutions have limitations in generalization and variability. Sub-band-based approaches simplify complexity and temporal information but yield similar results.

Table 3. Comparative analysis of speech recognition methodologies

Sr. No.	Speech Recognition Approaches	Advantages	Disadvantages
1	Acoustic Phonetic Recognition	It speeds up the processing of related words.	1. Due to the lengthy execution time of each isolated word, it is not frequently employed in commercial applications.
2	Pattern Recognition Approach	Word-to-word matching will occur, making pattern recognition rapid, simple, and automatic.	It helps with word-to-word matches. The key issue is template storage and the lengthy process. If there is a new variant of the pattern, it fails to recognize speech.
3	Template Based Approach	Discrete words benefit from it more. The segmentation and classification of small variable units result in fewer errors.	Expensive due to the vast vocabulary size and reference templates for each term. The preparation and template matching processes take additional time.
4	Dynamic Time Wrapping	Continuity is less crucial because it can fill up gaps in the sequence with incomplete data. Consistent time synchronization between the reference and test pattern.	It matches two provided sequences within specific bounds. It takes the longest to perform complex computational operations.
5	Vector Quantization Approach	1. It helps with effective data reduction.	1. It helps with effective data reduction.
6	Statistical Based Approaches (Hidden Markov Method)	The HMM has a vast vocabulary and can train with much data. It has a precise mathematical foundation.	Increasing computing complexity significantly Requires many data.
7	Artificial Approach Neural Network	1. It can efficiently complete time-consuming computing operations. It can change from the initial training model without making mistakes and automatically train the data and teach the system. It can also handle noisy, low-quality data well.	1. It produces vocabulary. Ineffective results with a broad It costs a lot since training involves numerous iterations over substantial training data. It takes more time to train. Due to the complex structure of the neural network, errors happen more often.
8	Neural network-based recognition speech	Unlike recurrent neural network-based speech improvement algorithms, neural network-based speech improvement algorithms are significantly faster.	Neural networks' "black box" character, increased computing load, propensity for overfitting, and the empirical nature of model building.
9	Fuzzy Logic based on recognition speech	Even if precise inputs are not necessary, it is robust. It is simple to modify to enhance the system's performance.	It is challenging to set precise fuzzy rules and membership functions. A fuzzy knowledge-based system needs a thorough testing approach for validation and verification.
10	Wavelet-based speech recognition	Wavelets provide localization in time and frequency. Wavelets use level-dependent thresholds to help eliminate additive sounds, which are present at all levels. Wavelet transform is a good alternative for pith detection because it can be used with non-stationary signals.	A lack of phase information, inadequate directionality, and shift sensitivity.
11	Optimization algorithm-based speech recognition	It resolves computation-related issues.	Inability variability to generalize and difficulty handling
12	Sub-band-based recognition speech	It produces comparable results. It is durable.	Both complexity and temporal information are lost.

These approaches have improved significant advancements in speech recognition technology, enabling its widespread use in various applications, such as virtual assistants, transcription services, and voice-activated systems.

An overview of various studies and research projects in the area of speech recognition and dialect identification is given in the following table. This research uses Hidden Markov Models (HMM), Mel-Frequency Cepstral Coefficients (MFCC), and various machine learning algorithms spanning various languages and databases. The research's varying recognition rates and accuracy levels highlight the difficulties and advancements in the field. Notably, studies on dialect identification for languages including Kannada, German, Arabic, and Assamese have yielded accuracy percentages between 62.33% and 85.7%. For languages like Bengali, Hindi, Telugu, Urdu, Assamese, Punjabi, and Manipuri, language recognition systems have also been created; their detection accuracy ranges from 46.67% to 86.67%. These works show how speech recognition and dialect identification have many uses in linguistics and technology.

Table 4. Comparative analysis of speech recognition technologies for continuous speech

Sr. No.	Speech Recognition System	Language	Database Used	Database Size	Techniques and Toolkits/Systems	Recognition Rate
1	Phonetic transcription [34]	English	DARPA	3267 Sentences	Hidden Markov Model (HMM)	85%
2	Speech Recognition [35]	English	Developed during research	1000 Words	HMM on Intel 486 PC	Word Error Rate (WER) 3.4%
3	Speech Recognition for Korean Language [36]	Korean	Developed during research	244 Words and 5610 Sentences	HMM	89%
4	HMM-based Speech Recognition [37]	Northern American	TIMIT	3648 training utterances and 1344 test utterances	MFCC with NICO	63.5%
5	Speech Recognition [38]	British English	BNC2A-J LOB1	1000 Sentences or 16522 Words	PCMA	WER reduced by 2.5%
6	Across-word Model Speech Recognition [39]	English	VERBMOBIL II	100,000 words	HMM	WER reduced by 2.1%
7	Polysyllabic Word Speech Recognition [40]	Dutch	VIOS	1463 Polysyllabic words in 885 utterances	HMM with HTK	81.1%
8	GiniSVM Speech Recognition [41]	English	Developed during research	36 Words (26 letters and 10 digits) 46730 training and 3112 testing utterances	GiniSVM	WER reduced by 5%
9	Speech Recognition for Indian Language [42]	Tamil, Telugu, and Marathi	Developed during research	Marathi-155541 Sentences Tamil-303537 Sentences Telugu-444292 Sentences	HMM with Sphinx – 2	WER for Marathi-23.2% Tamil- 20.2% Telugu-28%
10	Large Vocabulary Speech Recognition [43]	English	WSJCAM0	50000 Words	MFCC	WER reduced by 3.2%
11	Sparse Code Speech Recognition [44]	English	TIDIGITS	8623 utterances and contains 28,329 words	HMM with HTK	WER achieved up to 15%
12	Speech Recognition for Arabic Language [45]	Arabian	Developed during research	415 Sentences	HMM with Sphinx and HTK	WER is 5.78% and 5.45% with and without diacritical marks
13	Accents Detection for Indian English [46]	Marathi, Arabic Language	Developed during research	20 speakers from Iraq who speak Arabic as their mother tongue and who speak Marathi as their mother tongue	Acoustic feature approach	Speaker with Marathi accent 92.18% Speaker with Arabic Accent 93.21 for formant frequency (f0) respectively
14	Automatic Dialect Detection for Spontaneous Speech [47]	Latin American Spanish	Spanish-speaking countries in Latin America have three distinct accents: Cuba, Peru, and Puerto Rico (PR)	109 speakers (male and female) spontaneous Speech of Cuba, Peru, and Puerto Rico (PR) dialects	Universal background model (UBM), perceptual Minimum Variance Distortionless Response (PMVDR), shifted delta cepstrum (SDC)	MFCC + UBM 57.9, 68.4, 73.6 for 0dB (worst noise condition), 10dB (Gaussian white noise), and clean condition, respectively. PMVDR + SDC + UBM 76.3, 79, 79 for 0dB (worst noise condition), 10dB (Gaussian white noise) and clean condition, respectively.
15	Impact of Hindi Language on Other Indian Languages with Accent and Features [12]	Hindi, Marathi, Marwadi and Urdu	Developed during research	900 speech signals from 18 male and 12 female speakers of Hindi, Marathi, Marwadi, and	MFCC and LPC with Acoustic Features	Hindi-71.15% Marwadi-70.64% Marathi-71.91% Urdu-69.82%

				Urdu		
16	Identification of Kannada Dialects [17]	Kannada Language	Vowel dataset of Native speakers of Kannada	The remaining 20% (520 vowels) of the testing set were predicted using the training set (2080 vowels).	Random forest (RF), Extreme Random forest (ERF), and extreme gradient boosting (XGB) algorithms	77.22% and 74.13% with random forest for SV (simple validation) and CV (cross-fold validation), respectively, and 62.33%, 64.96% for ERF (extreme random forest), and 75.56%, 73.21% for XGB
17	Identification of German Dialects [18]	German Language	The dataset contained speakers from Basel, Bern, Lucerne, and Zurich.	Nearly 25,000 instances in the dataset were split into training, development, and test partitions.	Classifier Ensembles	The accuracy of the f1 score for the random baseline and SVM ensemble is 0.25 and 0.62, respectively.
18	Identify Arabic Utterance Dialects using Audio and Text Transcriptions [19]	Arabic Dialects in South Africa	Arabic dialect dataset, which contains audio and ASR transcripts videos	ASR transcripts of broadcast, debate, and discussion programs from videos.	ASR Transcripts and i-vector	Arabic dialects are utilizing ASR ranscripts and iVectors. A meta-classifier produced the best results with an accuracy rate of 71.7%.
19	Identification of Assamese Language Dialects using Spectral Features [7]	Assamese Language	Newly created corpus by researchers	Twenty-two triers (4 hours and 22 minutes each) served as angular bandpass filters for Assamese language analysis on 38 speakers, totaling 6 hours and 8 minutes. Additionally, information was gathered over three hours on the correlation between speech frequency (f) and 30 Kamrupi dialect speakers and 27 Goalparia dialect speakers.	Gaussian Mixture Model and Gaussian Mixture Model with a Universal background	Accuracy of 85.7% for GMM
20	Acoustic Features Based Word Level Dialect Classification [8]	British English	Nine corpora of nine distinct regions of the British Isles	speech corpus IViE (Intonational Variation in English)	SVM and Ensemble Methods	accuracy of 80.5% and 78.8% using simple and 5-fold cross-validation, respectively.
21	Development of a Language Identification System [9]	Bengali, Hindi, Telugu, Urdu, Assamese, Punjabi, and Manipuri	Utilized speech utterances from these seven languages	Utilized 15 utterances per language, each lasting 10-60 seconds	Employed a Neural Network (NN) classifier	Achieved identification accuracies as follows: Bengali- 86.67%, Hindi-60%, Telugu-86.67%, Urdu-73.33%, Assamese-46.67%, Punjabi- 80%, Manipuri-53.33%.
22	Automated Recognition of Dialects and Accents in South Indian English [11]	Telugu, Kannada, Tamil, Marathi Speakers from the Southern Region	Utilized a corpus of spoken South Indian English by Telugu, Marathi, Kannada, and Tamil speakers	Employed 4 test utterances for evaluation	Utilized Mel-Frequency Cepstral Coefficients (MFCC) and Hidden Markov Models (HMM)	Achieved an 80% improvement using the MFCC subband method.
23	Development of Acoustic-Phonetic Features for Dialect Identification in Hindi Speech [21]	Various Dialects of Hindi	Utilized a corpus of Hindi dialects spoken by native speakers	Included a database of 300 sentences for analysis	Employed Mel-Frequency Cepstral Coefficients (MFCC) and Perceptual Linear Prediction coefficients (PLP)	Achieved identification accuracies of 71% with MFCC, 68% with PLP, and 72% with MF-PLP.
24	Create corpora for the Marathi language and explore the use of the Sphinx engine for automatic speech recognition[10]	Marathi	Newly created corpus by Researchers	Marathi dictionary of 10,000 words.	SPHINX (Speech Recognizer Tool)	

5. APPLICATIONS

Understanding speech signal processing from a language aspect opens up various exciting applications that leverage the insights gained from analyzing spoken language. Here are some potential applications.

- Automatic Speech Recognition (ASR): ASR enhances transcription accuracy and dialect detection by considering speech patterns, accents, and dialects alongside context and co-articulation effects.
- Language Dialect Identification (LDI): LDI systems support linguistic forensics, telemedicine, education, entertainment, and customer service, facilitating cultural segmentation, marketing insights, behavior prediction, and customized product development.
- Speech-to-Text Captioning and Subtitling: Understanding speech signal processing enables automatic captions and subtitles for video accessibility and language translation, promoting inclusivity and extending content reach.
- Voice Assistants and Chatbots: Language-aware speech signal processing improves voice assistant interactions by enhancing query understanding, context, and user intent.
- Speaker Diarization: Accurate speaker diarization benefits from language clarity in separating and naming voices in audio recordings.
- Emotion and Sentiment Analysis: Linguistic speech signal processing aids in emotion and sentiment analysis for market research, customer service, and mental health analysis.
- Language Learning and Pronunciation Improvement: Real-time feedback on pronunciation, intonation, and accent assists language learners in improving their speaking abilities.
- Voice Analytics in Call Centers: Spoken language analysis enhances customer service and agent training for more effective client interactions.
- Accent Conversion and Synthesis: Accent conversion and synthesis offer personalization and accessibility in voice-based services and language learning by allowing users to select or learn different accents.
- Forensic Speaker Identification: Speech signal processing aids forensic investigators in identifying speakers from audio evidence for criminal case details.
- Speech-Based Content Indexing and Search: Linguistic understanding of spoken audio improves indexing for audio databases, simplifying searches for specific themes or phrases in voice recordings.
- Language Assessment and Disorders: Speech signal processing helps speech therapists diagnose and monitor language disorders and speech problems.
- Sentiment-Based Marketing and Advertising: Analyzing sentiment in spoken language helps businesses tailor marketing and advertising strategies to align with their target audience's emotions and preferences.

These applications demonstrate how integrating Speech signal processing with language understanding can lead to significant advancements across various industries, enriching human-computer interactions and communication processes.

6. DISCUSSION

This paper has discussed the review of spoken language dialects for various Indian and non-Indian languages with different techniques and corpora used to recognize the dialect and also focused stages involved for language dialect identification through speech signal processing. This paper also highlights how linguistic knowledge aids in discriminating between various languages and their dialects. By examining acoustic variables such as waveform, spectral (including MFCC, SDCC, and more), and prosodic elements, which reflect tone and speaking style, the survey investigates the combination of speech signal processing and linguistics to improve dialect identification. These traits make it easier to identify dialects and support linguists in their research on language diversity and evolution.

A detailed explanation of the processes of dialect identification, including speech signal pre-processing, which uses a variety of mathematical and computer approaches to analyze, manipulate, and interpret speech signals, is provided in the following portion of the study. The goal is to get useful information from the speech waveform, which can be used for automatic speech recognition, identifying the speaker, and figuring out what dialect they speak.

In order to improve signal strength and distinguish between various speech signals, feature extraction seeks to minimize the dimensionality of speech signal vectors. The following section investigates strategies for keeping important information while removing extraneous features, including MFCC, LPCC, and PCA.

The paper also discussed different ways to recognize speech, such as the Acoustic Phonetic approach, the Pattern recognition approach, etc. At the end of the paper, corpora, speech recognition techniques, database size, and recognition rate are looked at to see how well-existing speech recognition systems work with continuous speech. This research paper aims to deepen the understanding of speech signal processing from a language perspective, explore the linguistic features embedded in speech signals, and contribute to advancing speech technology and its applications, like speech recognition and spoken language dialect identification.

7. CONCLUSION

This paper has presented a literature review of spoken language dialect identification (LDI) and classification for Indian and non-Indian languages, along with how linguistic knowledge in Speech signal processing offers a thorough and precise method for identifying dialects. Significant LDI work has been done for non-Indian languages, whereas LDI work is in progress for Indian languages. Recognizing spoken language dialects becomes very difficult due to issues such as the unavailability of the annotated corpus, the use of multilingual words in speaking, and the regional impact of the area. Accurately distinguishing between closely related dialects that share many linguistic similarities is a challenging task in language dialect identification for Indian languages, and it becomes even more challenging for the Marathi language due to its vast set of dialects. It is evident from the review that many authors have implemented LDI systems using linguistic, ML, or hybrid approaches. Multiple ML techniques or a combination of speech signal processing and

ML techniques are used for comparing results. It is also noticed that combining linguistic expertise with advanced processing techniques makes it possible to create more effective models for distinguishing dialect variations based on

speech patterns. Less work on LDI is reported for Indian languages like Marathi and Gujarati. Development of appropriate techniques and methods of LDI for such languages is necessary.

In the future, this work will highlight the state of spoken language dialect identification (LDI) for non-Indian and Indian languages, highlighting the difficulties and developments in the area. Significant obstacles include the lack of annotated corpora, the complexities of multilingual speech, and the regional influence on dialects, particularly in languages like Gujarati and Marathi. Though there is a lot of LDI research on non-Indian languages, as the literature review shows, there has not been much advancement for Indian languages, especially Marathi and Gujarati.

The research will continue to focus on Marathi and Gujarati, develop annotated corpora for Indian languages, handle multilingual issues, investigate regional influences, advance machine learning techniques, and emphasize hybrid approaches. Furthermore, the creation and application of strong dialect identification models in real-world contexts depend heavily on promoting interdisciplinary cooperation and closing the gap between scholarly study and industry implementations. In general, the future mentioned above is intended to further the field of spoken language processing and LDI research by offering solutions to the particular difficulties presented by Indian languages.

8. REFERENCES

[1] J. K. Chambers and P. Trudgill, "Dialectology," chapter one, pp. 4-9, 2nd edition, Cambridge University Press, 1998.

[2] H. Li, B. Ma, and K. A. Lee, "Spoken Language Recognition: From Fundamentals to Practice," Proceedings of the IEEE, vol. 101, pp. 1136- 1159, 2013.

[3] J. Lee, K. Kim, and M. Chung, "Korean Dialect Identification Using Intonation Features," no. April, 2021, doi: 10.13140/RG.2.2.19397.17126.

[4] S. B. Patil, N. V. Patil, and A. S. Patil, "Speaker Independent Isolated Word Recognition using HTK for Varhadi a Dialect of Marathi," Int. J. Eng. Adv. Technol., vol. 9, no. 3, pp. 748–751, 2020, doi: 10.35940/ijeat.b3832.029320.

[5] M. Nanmalar, P. Vijayalakshmi, and T. Nagarajan, "Literary and Colloquial Dialect Identification for Tamil using Acoustic Features," IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON, vol. 2019- Octob, pp. 1303–1306, 2019, doi 10.1109/TENCON.2019.8929499.

[6] H. S. Das and P. Roy, "Optimal prosodic feature extraction and classification in parametric excitation source information for Indian language identification using neural network based Q-learning algorithm," Int. J. Speech Technol., vol. 22, no. 1, pp. 67–77, 2019, doi: 10.1007/s10772-018-09582-6.

[7] T. Ismail and L. Joyprakash Singh, “Dialect Identification of Assamese Language using Spectral Features,” Indian J. Sci. Technol., vol. 10, no. 20, pp. 1–7, 2017, doi: 10.17485/ijst/2017/v10i20/115033.

[8] N. B. Chittaragi and S. G. Koolagudi, "Acoustic features based word level dialect classification using SVM and ensemble methods," 2017 10th Int. Conf. Contemp. Comput. IC3 2017, vol. 2018-Janua, no. January, pp. 1–6, 2018, doi: 10.1109/IC3.2017.8284315.

[9] C. Madhu, A. George, and L. Mary, "Automatic language identification for seven Indian languages using higher level features," 2017 IEEE Int. Conf. Signal Process. Informatics, Commun. Energy Syst. SPICES 2017, 2017, doi: 10.1109/SPICES.2017.8091332.

[10] A. Ankit et al., "Acoustic Speech Recognition for Marathi Language Using Sphinx," ICTACT J. Commun. Technol., vol. 7, no. 3, pp. 1361– 1365, 2016, doi: 10.21917/ijct.2016.0201.

[11] Ashok Shigli, Ibrahim Patel and K Srinivasa Rao, "Automatic Dialect and Accent Speech Recognition of South Indian English,"

International Journal of Latest Trends in Engineering and Technology, pp 103-111,2016.

[12] R. H. Chaudhari, K. Waghmare, and B. W. Gawali, "Accent Recognition using MFCC and LPC with Acoustic Features," Int. J. Innov. Res. Comput. Commun. Eng., vol. 3, no. 3, pp. 2128–2134, 2015.

[13] G. Liu and J. H. L. Hansen, "A systematic strategy for robust automatic dialect identification," Eur. Signal Process. Conf., no. Eusipco, pp. 2138–2141, 2011.

[14] R. W. M. Ng, T. Lee, C. C. Leung, B. Ma, and H. Li, "Analysis and selection of prosodic features for language identification," 2009 Int. Conf. Asian Lang. Process. Recent Adv. Asian Lang. Process. IALP 2009, no. Figure 2, pp. 123–128, 2009, doi: 10.1109/IALP.2009.34.

[15] A. Etman and A. A. L. Beex, "Language and Dialect Identification: A survey," IntelliSys 2015 - Proc. 2015.

[16] SAI Intell. Syst. Conf., pp. 220– 231, 2015, doi: 10.1109/IntelliSys.2015.7361147.

[17] D. Raval, V. Pathak, M. Patel, and B. Bhatt, "End-to-End Automatic Speech Recognition for Gujarati," Proc. 17th Int. Conf. Nat. Lang. Process., pp. 409–419, 2020.

[18] [17] N. B. Chittaragi and S. G. Koolagudi, "Acoustic-phonetic feature based Kannada dialect identification from vowel sounds," Int. J. Speech Technol., vol. 22, no. 4, pp. 1099–1113, 2019, doi: 10.1007/s10772- 019-09646-1.

[19] A. M. Ciobanu, S. Malmasi, and L. P. Dinu, "German Dialect Identification Using Classifier Ensembles," COLING 2018 - 27th Int. Conf. Comput. Linguist. Proc. 5th work. NLP Similar Lang. Var. Dialects, VarDial 2018, pp. 288–294, 2018.

[20] S. Malmasi and M. Zampieri, "Arabic dialect identification using iVectors and ASR transcripts," VarDial 2017 - 4th Work. NLP Similar Lang. Var. Dialects, Proc., no. 2015, pp. 178–183, 2017, doi: 10.18653/v1/w17-1222.

[21] S. Sinha, A. Jain, and S. S. Agrawal, "Fusion of multi-stream speech features for dialect classification," CSI Trans. ICT, vol. 2, no. 4, pp. 243–252, 2015, doi: 10.1007/s40012-015-0063-y.

[22] S. Sinha, A. Jain, and S. S. Agrawal, "Acoustic-phonetic feature based dialect identification in Hindi speech," Int. J. Smart Sens. Intell. Syst., vol. 8, no. 1, pp. 235–254, 2015, doi: 10.21307/ijssis-2017-757.

[23] A. Ankit et al., "Acoustic Speech Recognition for Marathi Language Using Sphinx," ICTACT J. Commun. Technol.,

vol. 7, no. 3, pp. 1361– 1365, 2016, doi: 10.21917/ijct.2016.0201.

[24] Caka,Nebi.(2015).https://www.researchgate.net/post/What-are-the- Spectral-and-Temporal-Features-in-Speech-signal/54fb90d1d11b8b897b8b4567

[25] S. Poornima, "Basic Characteristics of Speech Signal Analysis," Int. J. Innov. Res. Dev., vol. 5, no. 4, pp. 1–5, 2016.

[26] Essien,Akpan.(2015).https://www.researchgate.net/post/What-are-the-Spectral-and-Temporal-Features-in Speech- signal/550747c7d11b8b5d358b4630

[27] B. Ma and H. Li, "A Comparative Study of Four Language Identification Systems," Int. J. Comput. Linguist. Chinese Lang. Process. Vol. 11, No. 2, June 2006, vol. 11, no. 2, pp. 159–182, 2006.

[28] Z. Tang, D. Wang, Y. Chen, L. Li, and A. Abel, "Phonetic temporal neural model for language identification," IEEE/ACMTransactions on Audio, Speech, and Language Processing, vol. 26, no. 1, pp. 134–144, 2018.

[29] M. Tirusha, "Multilingual {Phonetic} {Features} for {Indian} {Language} {Identification}," no. February, 2020.

[30] Kamble, B. C., "Speech Recognition Using Artificial Neural Network–A Review," Int. J. Comput. Commun. Instrum. Eng, 3(1), 61-64, 2016.

[31] A. Pangotra, "Review on Speech Signal Processing & Its Techniques," Eur. J. Mol. Clin. Med., vol. 7, no. 7, pp. 3049–3052, 2020.

[32] S. Dass and A. K. Yadav, "Comparative Analysis of Speech Processing Techniques at Different Stages," Springer Int. Conf. Pattern Recognit. Tech., no. 1, 2017.

[33] M. K. Sharma, "Speech Recognition: a Review," Int. J. Adv. Eng. Res. Dev., vol. 1, no. 12, 2014, doi: 10.21090/ijaerd.011244.7.

[34] V. Waghmare Shri Swami Vivekanand Shikshan Sanstha, P. K. Kurzekar, R. R. Deshmukh, V. B. Waghmare, and P. P. Shrishrimal, "Continuous Speech Recognition System: a Review," Asian J. Comput. Sci. Inf. Technol. J. homepage, vol. 4, no. 6, pp. 62–66, 2014, doi: 10.15520/ajcsit.v4i6.3.

[35] S. E. Leninson, A. Ljolie, L.G. Miller, "Continuous Speech Recognition from a Phonetic Transcription," Acoustics, Speech, and Signal Processing, vol.1 pp. 93 – 96, Apr 1990.

[36] Paul Bamberg, Yen-lu Chow, Laurence Gillick, Robert Roth and Dean Sturtevant, "The Dragon Continuous Speech Recognition System: A Real-Time Implementation," Proceedings of DARPA Speech and Natural Language Workshop, Hidden Valley, Pennsylvania, pp. 78-81, June 1990.

[37] H. R. Kim, K. W. Hwang, N. Y. Han, and Y. M. Ahn, "Korean Continuous Speech Recognition System Using Context-Dependent Phone SCHMMs," Proceedings of the Fifth Australian International Conference on Speech Science and Technology, vol. II, pp.694 - 699, 1994.

[38] Simon King, Paul Taylor, "Detection of phonological features in continuous speech using neural networks," Computer Speech & Language, vol. 14, Issue 4, October 2000, pp. 333–353.

[39] Mark Huckvale and Alex Chengyu Fang, "Using Phonologically- Constrained Morphological Analysis in Continuous Speech Recognition," Computer Speech and Language, vol. 16, pp.165-181, 2002.

[40] Achim Sixtus and Hermann Ney, "From within-word model search to across-word model search in large vocabulary continuous speech recognition," Computer Speech and Language, Vol 16, 2002, pp.245– 27.

[41] Odette Scharenborg, Louis ten Bosch, Lou Boves, "Early recognition of polysyllabic words in continuous speech," Computer Speech and Language, Vol 21, pp. 54–71, 2007.

[42] Veera Venkataramani, Shantanu Chakrabartty, William Byrne "Ginisupport vector machines for segmental minimum Bayes risk decoding of continuous speech," Computer Speech and Language, Vol 21, pp. 423–442, 2007.

[43] Gopalakrishnan Anumanchipalli, Rahul Chitturi, Sachin Joshi, Rohit Kumar, Satinder Pal Singh R.N.V. Sitaram, S P Kishore, "Development of Indian Language Speech Databases for Large Vocabulary Speech Recognition Systems," International Institute of Information Technology, Hyderabad, India July 2007.

[44] Giulia Garau, Steve Renals "Combining Spectral Representations for Large-Vocabulary Continuous Speech Recognition," IEEE Transactions on Audio, Speech, and Language Processing, vol. 16, no. 3, March 2008.

[45] W.J. Smit, E. Barnard, "Continuous speech recognition with sparse coding," Computer Speech and Language, vol 23, pp. 200–219, 2009.

[46] Mohammad Abushariah, Raja Ainon, Roziati Zainuddin, Moustafa Elshafei, and Othman Khalifa, "Arabic Speaker-Independent Continuous Automatic Speech Recognition Based on a Phonetically Rich and Balanced Speech Corpus," The International Arab Journal of Information Technology, vol. 9, No. 1, January 2012.

[47] N. B. Chittaragi and S. G. Koolagudi, "Automatic dialect identification system for Kannada language using single and ensemble SVM algorithms," vol. 54, no. 2. Springer Netherlands, 2020.