Survey on Automatic Speech Recognition Systems for Indic Languages

Nandini Sethi(✉) and Amita Dev

Indira Gandhi Delhi Technical University for Women, Delhi 110006, India
{nandini053phd20,vc}@igdtuw.ac.in

Abstract. For the past few decades, Automatic Speech Recognition (ASR) has gained a wide range of interest among researchers. From just identifying the digits for a single speaker to authenticating the speaker has a long history of improvisations and experiments. Human's Speech Recognition has been fascinating problem amongst speech and natural language processing researchers. Speech is the utmost vital and indispensable way of transferring information amongst the human beings. Numerous research works have been equipped in the field of speech processing and recognition in the last few decades. Accordingly, a review of various speech recognition approaches and techniques suitable for text identification from speech is conversed in this survey. The chief inspiration of this review is to discover the prevailing speech recognition approaches and techniques in such a way that the researchers of this field can incorporate entirely the essential parameters in their speech recognition system which helps in overcoming the limitations of existing systems. In this review, various challenges involved in speech recognition process are discussed and what can be the future directives for the researchers of this field is also discussed. The typical speech recognition trials were considered to determine which metrics should be involved in the system and which can be disregarded.

Keywords: Speech recognition · Acoustic modelling · Hidden Markov model · Dynamic time wrapping · Mel-frequency Cepstrum Coefficient

1 Introduction

Human voice has been the major mode of communication, interacting with machines has evolved a lot from identifying digits to the complex Automatic Speech Recognition (ASR) till date. The desire to automate simple tasks to complex tasks has necessitated human-machine interactions. Over the past decades, a lot of research has been carried out in order to create an ideal system which can understand and analyse continuous speech in real time and perform tasks accordingly. Some of which are Speech-to-text conversions, biometric identifications, home automation and has also highly benefited disabled persons. Advancements in the deep neural networks has made it all possible. Hidden Markov Model (HMM) hydride with Deep neural networks (DNN) and Recurrent Neural Networks has achieved remarkable performance in many large vocabulary speech recognition tasks [42].

But this was not as easy as what we see today. In terms of evolution, it can be organised and shown in Table 1.

A. Dev et al. (Eds.): AIST 2021, CCIS 1546, pp. 85–98, 2022.
https://doi.org/10.1007/978-3-030-95711-7_8

Table 1. Evolution of Speech Recognition Systems

Generation	Technology	Timeline
First generation	First attempt	1950s to 1960s
Second generation	Template based technology	1960s to 1970s
Third generation	Statistical modelling	1980s and 2000
Fourth generation	Advancements in deep Neural networks	after late 2000

1.1 First Generation

Earlier attempts in the field of ASR were made between 1950s to 1960s, when researchers were experimenting with the fundamental ideas of acoustic phonetics. In 1952, at Bell Laboratories in USA, David, Balashek and Biddulph built a system which could recognize digits for a single isolated speaker using formant frequencies measured during vowel regions of each digit. Further, at University College in England, Fry and Denies built a system which could recognize four vowels and nine consonants. This was the benchmark at that point of time in recognizing phonemes with much better accuracy as before. In the 1960s computers were not fast enough which was the limitation for the hardware. Other than this non-uniformity of time scales in speech was also a hurdle. To overcome this problem Martin and his colleagues at RCA labs developed a set of elementary time-normalised methods. This helped in reliably recognizing the start and end of a speech that reduced the variability of the recognition scores.

1.2 Second Generation

During the late 1960s and 1970s, ASR achieved many benchmarking milestones. Dynamic programming methods or Dynamic Time Wrapping (DTW) was introduced which helped in aligning a pair of speech utterances and also algorithms for connected word recognition. Many attempts were made during this time, for example, by IBM labs, AT & T Bell Labs, DARPA program and many more.

1.3 Third Generation

During the late 1980s, the focus was on building a more robust system which could recognize a fluently spoken string. A wide range of algorithms and experiments were performed to obtain a concatenated string of different words spoken. One of the key technologies developed during this time was the Hidden Markov Model (HMM). This technique was then boosted and was widely applied in every speech recognition research laboratory. Also, the idea of Neural networks was reintroduced in speech recognition. However, the concept was introduced earlier in the 1950s but was not useful because of practical limitations [44]. In the 1990s, after neural networks were reintroduced, many new innovations came in the area of pattern recognition.

1.4 Fourth Generation

For the past two decades, after so many successful attempts of improvising speech recognition systems, Deep Neural networks took speech recognition systems from just experimenting it on desk to some real-world applications for users. People can now interact and can make many tasks done just through voice command. For example, Ok Google, Siri, Alexa.

2 Speech Utterances

Based on the types of utterances, different speech recognition approaches are categorized into various groups in the way that they are capable to identify. Various types of speech utterances include Isolated speech, connected words, Continuous speech and spontaneous speech are shown in Fig. 1 and discussed below:

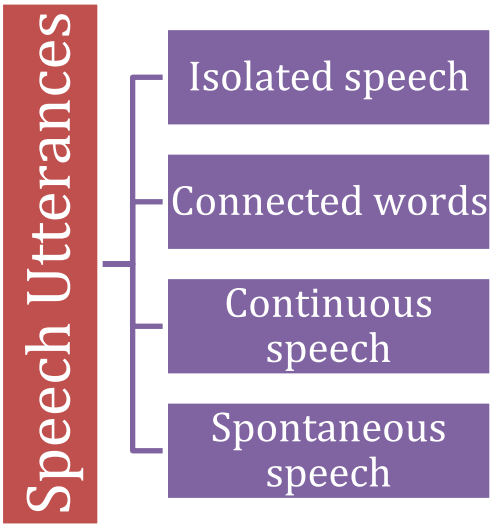

Fig. 1. Type of utterances

2.1 Isolated Speech

The recognizers which work with isolated speech requires every word to have noiselessness such as the absence of an audial signal on every side of the trial window. It accepts individual utterances at a specific point of time. This procedure includes two states named “Listen and Not-Listen”, in which the user is required to pause among the words consistently carrying out the processing during the period of pauses. It can also be termed as Isolated Utterance.

2.2 Connected Words

In the case of Connected word, it requires a least gap among utterances to permit the flow of speech smoothly. These type of speech utterances are slightly similar to isolated speech.

2.3 Continuous Speech

The recognizers which work with Continuous speech allows the operators to speak nearly in a natural way, whereas the processor chooses the context. Primarily, it characterizes the computer transcription. This type of Recognizers which work with the continuous speech are supplementary hard to produce as they implement exclusive procedures to choose on the utterance boundaries.

2.4 Spontaneous Speech

Spontaneous speech is a kind of speech which can be considered as a natural speech, not as the trained one. An Automatic Speech Recognition system with this type of speech dimensions has to be capable to identify the owner of normal speech characters such as utterances which work altogether for instance the “ums” and “ahs”, involves the minor stammers.

3 Speech Recognition Overview

Automatic Speech Recognition or ASR, is the procedure that permits humans to utilize their speeches to communicate with a system in such a way that, in its utmost cultured distinctions, be similar to natural conversation of humans. It can be divided into five different components shown in Fig. 2 and discussed as follows:

3.1 Pre-processing

In this step, some basic functions are performed before extracting any features. For Example, noise removal, endpoint detection, pre-emphasis and normalisation.

3.2 Feature Extraction

Features which will be used to differentiate between different phonemes and eventually to words and sentences are extracted. Most commonly extracted feature for ASR is Mel frequency cepstral coefficients (MFCCs) [43]. Since the mid-1980s MFCCs are the most widely used feature in ASRs. Discrete Wavelet Transform (DWT), Wavelet Packet Transform (WPT), Linear prediction cepstral coefficients (LPCC) and many more features are available with their strengths and weaknesses which can be used as required [40].

3.3 Classification

Numerous approaches have been done in order to find an optimal classifier which could correctly recognize speech segments under various conditions. Some of the classification techniques used are Artificial Neural Networks (ANNs), Hidden Markov Model (HMM).

3.4 A Language Model

Contains the knowledge specific to a language. This model is required to recognise phonemes and eventually represent meaningful representations of the speech signal [41].

3.5 Acoustic Modelling

Acoustic modelling establishes a relationship between acoustic information and language construct in SR [35].

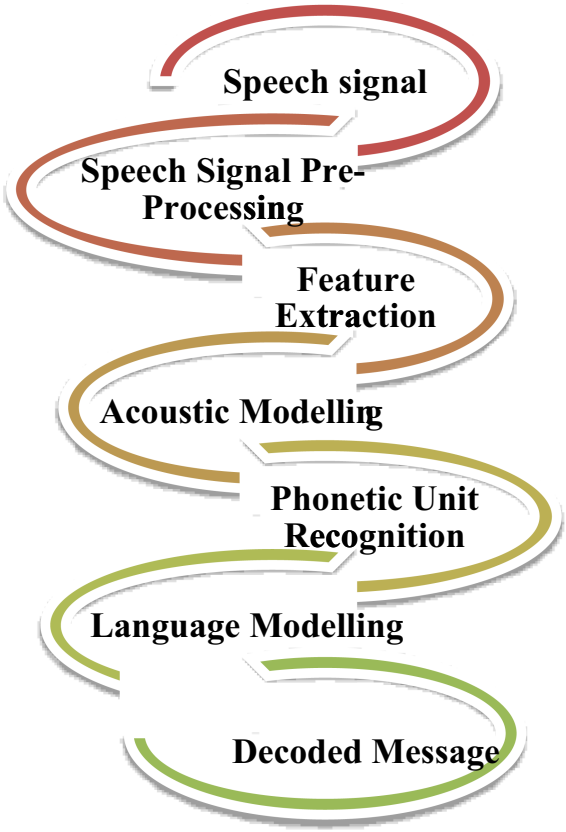

Fig. 2. Automatic speech recognition process

4 Speech Recognition Approaches

Speech Recognition has various techniques which can be further categorized into three major categories shown in Fig. 3 and discussed below:

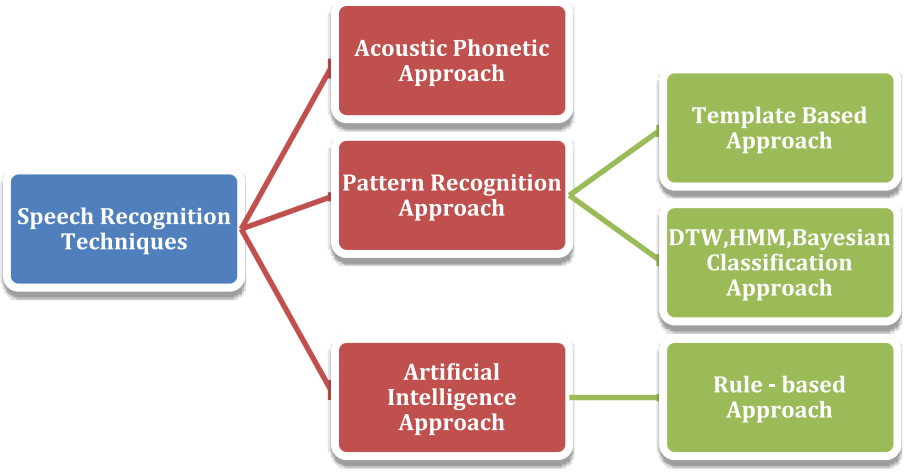

Fig. 3. ASR approaches

4.1 Acoustic Phonetic Approach

Acoustic phonetic approach, postulates that there exists a phoneme unit which is a building block of a speech and can be characterised in a set of acoustic properties. These properties are highly variable with respect to the speaker and the environment [36]. The very first step is speech spectral analysis followed by feature extraction which translates spectral dimensions into a conventional feature that defines the phonetics characteristics. After speech spectral analysis the segmentation and labelling of the speech signal is performed which generate isolated regions by segmenting the speech signal. Finally, it identifies the appropriate word from the produced sequences of phonetic labels. However, this approach is not widely used.

4.2 Pattern Recognition Approach

The pattern recognition approach is utilized to identify patterns grounded on convinced conditions which is used to categorize into various the classes. It involves various steps namely:

- feature measurements,
- pattern training,
- pattern classification and

- decision logic.

Various measurements are taken place to outline a test pattern on an input speech signal. Reference patterns are generated for each speech sound identified. Reference patterns can be generated with the help of speech templates or by using a statistical model such as HMM. The model can be functional to an utterance, a term or an idiom [46]. Finally, a comparison is performed among the unknown patterns and reference patterns in this pattern classification. And in Decision logic the identity of the unknown is determined. This approach is primarily used in ASR systems.

4.2.1 Template Based Approach

The essential thought of the Template based approach is elementary. A compendium of ideal patterns of speech is combined as reference patterns describing the vocabulary of candidate utterances. Afterwards, by toning the unnamed words with every reference pattern the recognition of word is performed and selects the kind of the finest matching form. All the words of the templates are configuring. One of the vibrant origins in this approach is to reach at a distinguishable sequence of speech frames for a word by means of firm averaging procedure and to be contingent on the arrangement of limited spectral distance metrics to evaluate and distinguish among the patterns [45]. An alternative vital concept is to implement a specific form of dynamic programming to briefly line up the patterns to reason for the alterations in the talking rates transversely the speakers as well across the repetitions of the term by the matching speaker.

Depending on the context, this approach handles the varying form of a nation-wide contour. It postulates an only sensible quantity of situations. This approach is considerable sturdier and long-range forecasters. The implementation of patterns is planned to detect and replicate the utmost significant syllable-level structures [39] of the outline without doing much smoothing. To depict the template form, this approach exploits the supple utterance arrangement. The benefit of using this approach is that, it can avoid the faults happened due to classification or segmentation of smaller adjustable components of illustration phonemes [38].

4.2.2 DTW, HMM Based Approach

In Dynamic time wrapping, various templates are used to represent every class which is to be recognized. To improve the speaker variability or the pronunciation modelling it is preferred to utilize two or more reference templates for each class. All through the process recognition, a gap among an experimental sequence of speech and class template is considered. The stretched and wrapped forms of the reference templates are also implemented in the gap calculation, to disregard the influence of duration discrepancy among the experimental sequence of speech and class template. The predictable word matches to the track over the model that reduces the total gap [41]. To improve the performance of dynamic time wrapping the number of class template variants can be increased and the wrapping constraints can be loosed but on the outflow of storing space and computational needs. Due to improved generalization features and lower memory

necessities, HMM based approach is more widely used instead of dynamic time wrapping approach in various state of the art systems.

4.3 Artificial Intelligence Approach

Artificial Intelligence approach is the combination of both the pattern recognition and acoustic phonetic approaches. Few scholars established knowledge base of acoustic phonetic features for speech recognition system which classifies the rules for sound of the speech [37]. The methods based on templates provides slight intuition about humanoid speech processing, nevertheless these procedures have been very efficient in the development of a diversity of automatic speech recognition systems. On the contrary, verbal and phonic works providing intuitions about humanoid speech processing. Though, this method had only fractional accomplishment due to the complexity in computing proficient information.

5 Various Speech Recognition System

This study efforts to examine entirely all the published works for automatic speech recognition of various Indic languages. Works that denotes to speech recognition system in Indic languages or associated investigation on the Indian Automatic Speech Recognition datasets variety, investigational and non-experimental have been encompassed in this review paper. The necessity of the critical study is to identify the position of the investigation on automatic speech recognition. Various automatic speech recognition research studies for Indian languages emphasised on:

- LPC (Linear Predictive Coding),
- MFCC (Mel-frequency Cepstrum Coefficient),
- RASTA (Relative Spectra Processing)
- ZCPA (Zero Crossing with Peak amplitude), and
- Dynamic Time Wrapping (DTW) features.

Various ASR systems with their features and accuracy are discussed in Table 2.

Table 2. Survey-based on feature extraction technique

Authors	Feature extracted	Language	Accuracy
Thasleema TM et al. [1]	LPC, wavelet packet decomposition method	Malayalam	34% with LPC
			74% with wavelet packet decomposition method
Sinha et al. [2]	MFCC, PLP	Hindi	–
Dutta and Sarma et al. [3]	MFCC, LPC	Assamese	–
Kaur and Singh et al. [4]	MFCC, PLP, PNCC	Punjabi	WER of MFCC-13.19%
			WER of PLP-16.28%
			WER of PNCC-16.28%
Kadyan et al. [5]	LDA, SAT, fMLLR and MLLT	Punjabi	WER of 17.53
Ventateswarlu et al. [6]	LPCC and MFCC	Telegu	3–4% higher for MFCC instead of PLP
Kumar et al. [7]	MFCC and PLP	Hindi	MFCC-(92.0%)
			PLP – (73.36%)
Bharali and Kalita et al. [8]	MFCC	Assamese	81%
Bhowmik et al. [9]	MFCC	Bengali	86.19%
Mohamed and Lajish et al. [10]	MFCC	Malayalam	80.74%
ChellaPriyadharshini et al. [11]	MFCC, LDA, MLLT, fMLLR	Tamil	WER reduction of 15%
Manjunath and Rao et al. [12]	Afs (Articulatory Features)	Kannada, Telugu, Bengali, and Odia	(PER) of 10.4%
Darekar and Dhande et al. [13]	Cepstral, NMF and MFCC	Marathi	–

Various classification techniques are available for automatic speech recognition such as HMM, GMM, RNN, SOM, DE-HMM, DE-GMM, MPE, MMI and MLE. Techniques used in various research works with features extracted and language used is discussed in Table 3.

Table 3. Survey based on classification techniques

Author	Classifier used	Features extracted	Language
Kurian and Balakrishnan et al. [14]	HMM	–	Malayalam
Paul et al. [15]	ANN	MLP and LPC	Bangla
Sarma et al. [16]	ANN and SOM	MLP	Assamese
Sukumar et al. [17]	ANN	DWT	Malayalam
Bhuvanagirir and Kopparapu [18]	ANN	N/A	Malayalam
Dutta and Sarma [3]	RNN	LPC and MFCC	Assamese
Das et al. [19]	HMM	–	Bengali
Sarma et al. [20]	RNN, SOM and PNN	DWT	Assamese
Kumar et al. [21]	HMM and GMM	MFCC	Bengali and Odia
Patil and Pardeshi [22]	HMM	N/A	Devanagari
Patil and Pardeshi [23]	HMM	MFCC	Marathi
Hemakumar and Punitha [24]	HMM	LPC and RCC	Kannada
Patil and Rao [25]	HMM	acoustic–phonetic features	Hindi, Marathi
Dua et al. [26]	HMM	MFCC, PLP, MMIE, MME and MPE	Hindi
Bhat et al. [27]	Bayesian and HMM	–	Kannada
Pulugundla et al. [28]	TDNN and BRMN	–	Tamil, Telugu and Gujarati
Dua et al. [29]	HMM, GMM, DE-HMM, DE-GMM, MPE, MMI and MLE	MFCC, GFCC	Hindi
Samudravijaya et al. [30]	HMM, n-gram language modelling and RNNLM	MFCC	Hindi
Fathima et al. [31]	TDNN	Phonetic Features	Gujrati
Pandey and Nathwani [32]	DNN, DNN-HMM, KWS	Spectral and Prosodic features	Hindi
Pal et al. [33]	SGMM with LDA, MLLT and SAT	MFCC, delta and double delta features	Bengali

(*continued*)

Table 3. (*continued*)

Author	Classifier used	Features extracted	Language
Patel et al. [34]	GMM_HMM, DNN-HMM, KWS	–	Manipuri

6 Challenges and Future Directions in Speech Recognition

Robustness of an Automatic Speech Recognition system is the capability of the system to effectively deal with diverse characteristics of inconsistency in the speech (input) signal. The accuracy of a speech recognition system can be evaluated by a number of eminent factors. The utmost perceptible ones are: speaker, pronunciation, region, speech rate, context, channel and environment variability. In the development of ASR systems, these thought-provoking factors must be taken care and efficient models to be formed to deliver virtuous recognition precision regardless of these variabilities. In advanced level, ASR system development requires the accessibility of procedures or algorithms for instinctive generation of expression lexicons, instinctive generation of linguistic models for novel tasks, instinctive algorithms for speech segmentation, algorithm for finest utterance verification-rejection, attaining or exceptional humanoid presentation on ASR tasks. Some of the challenges and future directives are discussed below:

- Several Automatic Speech Recognition systems has absence of huge speech corpus. To build such a huge corpus must include tonal information, dialectal and prosodic information to perform more analytical processing of information.
- Language such as Punjabi, Bodo and Dogari are tonal Indic languages. An examination required to be accomplished by means of vocal tract information and pitch information about these dialects and their successive languages.
- Additional chief problem with vernaculars is a discrepancy of dialectal statistics. Rare studies were performed on mining the dialectal information of Indic dialects. This required to be united with speech methodologies to diminish Word Error Rate.
- Various works in this field implemented the bottle neck features. Various speech databases established in Indic dialects are grounded on noise free situation. In future, researchers can develop noisy datasets and develop speech recognition system on these datasets by utilizing various pitch characteristics and robust approaches to enhance the performance of the system.
- By utilizing the optimisation algorithm on model metrics, an effort can be made to improve the acoustical features. Very rare works has been worked on optimizing or refining the features. Study in other dialects emphases on previously recognized techniques of feature extraction such as MFCC. A limited studies have utilized hybridisation techniques of feature extraction for the refinement of feature.

7 Conclusion

Speech recognition is a standout amongst the utmost enabling zones of machine information since people do an ordinary movement of speech recognition. In this survey, various speech recognition techniques and their works are reviewed and tabulated different features extracted and classifier used on the (input) speech signal. Prominently, three distinct factors such as, approach, features extracted and accuracy measure were taken care for comparison and studying the prevailing works. The comprehensive analysis accomplished in this study will give the attainment happened in the field of automatic speech recognition to further articulate the research notions to overcome the existing yardstick outcomes for the scholars. At last, some of the research challenges and future directives are also addressed to lead the further research in the same direction. In future, researchers can develop noisy datasets and develop speech recognition system on these datasets by utilizing various pitch characteristics and robust approaches to enhance the performance of the system.

References

1. Thasleema, T.M., Kabeer, V., Narayanan, N.K.: Malayalam vowel recognition based on linear predictive coding parameters and k-NN algorithm. In: Proceedings of international conference on computational intelligence and multimedia applications (ICCIMA 2007), pp. 361–365 (2007)
2. Sinha, S., Agrawal, S.S., Olsen, J.: Development of Hindi mobile communication text and speech corpus. In: Proceedings of O-COCODSA, pp. 30–35 (2011)
3. Dutta, K., Sarma, K.K.: Multiple feature extraction for RNN-based Assamese speech recognition for speech to text conversion application. In: Proceedings of the international conference on communications, devices and intelligent systems (CODIS), pp. 600–603 (2012)
4. Kaur, A., Singh, A.: Optimizing feature extraction techniques constituting phone-based modelling on connected words for Punjabi automatic speech recognition. In: Proceedings of the 2nd International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, India, pp. 2104–2108 (2016b)
5. Kadyan, V., Mantri, A., Aggarwal, R.K., Singh, A.: A comparative study of deep neural network-based Punjabi—ASR system. Int. J. Speech Technol. **22**(1), 111–119 (2018)
6. Venkateswarlu, R.L.K., Teja, R.R., Kumari, R.V.: Developing efficient speech recognition system for Telugu letter recognition. In: Proceedings of International Conference on Computing, Communication and Applications, pp. 1–6 (2012)
7. Kumar, A., Dua, M., Choudhary, A.: Implementation and performance evaluation of continuous Hindi speech recognition. In: Proceedings of International Conference on Electronics and Communication Systems (ICECS), pp. 1–5 (2014a)
8. Bharali, S.S., Kalita, S.K.: Speech recognition with reference to Assamese language using novel fusion technique. Int. J. Speech Technol. **21**(2), 251–263 (2018). https://doi.org/10.1007/s10772-018-9501-1
9. Bhowmik, T., Chowdhury, A., Mandal, S.K.D.: Deep neural network-based place and manner of articulation detection and classifcation for Bengali continuous speech. Procedia Comput. Sci. **125**, 895–901 (2018)
10. Mohamed, F.K., Lajish, V.L.: Nonlinear speech analysis and modeling for Malayalam vowel recognition. Procedia Comput. Sci. **93**, 676–682 (2016)

11. Chellapriyadharshini, M., Tofy, A., Srinivasa, R.K.M., Ramasubramanian, V.: Semi-supervised and active-learning scenarios: efficient acoustic model refinement for a low resource Indian language. In: Computer and Languages, pp. 1041–1045 (2018)
12. Manjunath, K.E., Sreenivasa Rao, K.: Improvement of phone recognition accuracy using articulatory features. Circ. Syst. Sig. Process. **37**(2), 704–728 (2017). https://doi.org/10.1007/s00034-017-0568-8
13. Darekar, R.V., Dhande, A.P.: Emotion recognition from Marathi speech database using adaptive artifcial neural network. Biol. Inspired Cognit. Archit. **23**, 35–42 (2018)
14. Kurian, C., Balakrishnan, K.: Speech recognition of Malayalam numbers. In: Proceedings of the World Congress on Nature and Biologically Inspired Computing, pp. 1475–1479 (2009)
15. Paul, A.K., Das, D., Kamal, M.: Bangla speech recognition system using LPC and ANN. In: Proceedings of the 7th International Conference on Advances in Pattern Recognition, pp. 171–174 (2009)
16. Sarma, B.D., Sarmah, P., Lalhminghlui, W., Prasanna, S.M.: Detection of Mizo tones. In: Proceedings of Sixteenth Annual Conference of the International Speech Communication Association, pp. 934–937 (2015)
17. Sukumar, A.R., Shah, A.F., Anto, P.B.: Isolated question words recognition from speech queries by using artifcial neural networks. In: Proceedings of international conference on computing communication and networking technologies, pp. 1–4 (2010)
18. Bhuvanagirir, K., Kopparapu, S.K.: Mixed language speech recognition without explicit identifcation of language. Am. J. Sig. Process. **2**(5), 92–97 (2012)
19. Das, B., Mandal, S., Mitra, P.: Bengali speech corpus for continuous automatic speech recognition system. In: Proceedings of the International Conference on Speech Database and Assessments, pp. 51–55 (2011)
20. Sarma, B.D., Sarma, M., Sarma, M., Prasanna, S.R.M.: Development of Assamese phonetic engine: some issues. In: Proceedings of the annual IEEE India Conference (INDICON), pp. 1–6 (2013)
21. Kumar, S.B.S., Rao, K.S., Pati, D.: Phonetic and prosodically rich transcribed speech corpus in Indian languages: Bengali and Odia. In: Proceedings of International Conference Oriental COCOSDA held Jointly with 2013 Conference on Asian Spoken Language Research and Evaluation (O-COCOSDA/CASLRE), pp. 1–5 (2013a)
22. Patil, P.P., Pardeshi, S.A.: Devnagari phoneme recognition system. In: Proceedings of the Fourth International Conference on Advances in Computing and Communications (ICACC), pp. 5–8 (2014b)
23. Patil, P.P., Pardeshi, S.A.: Marathi connected word speech recognition system. In: Proceedings of the First International Conference on Networks and Soft Computing (ICNSC), pp. 314–318 (2014a)
24. Hemakumar, G., Punitha, P.: Automatic segmentation of Kannada speech signal into syllables and sub-words: noised and noiseless signals. Int. J. Sci. Eng. Res. **5**(1), 1707–1711 (2014)
25. Patil, V.V., Rao, P.: Detection of phonemic aspiration for spoken Hindi pronunciation evaluation. J. Phon. **54**, 202–221 (2016)
26. Dua, M., Aggarwal, R.K., Biswas, M.: Discriminative training using heterogeneous feature vector for Hindi automatic speech recognition system. In: Proceedings of International Conference on Computer and Applications (ICCA), pp. 158–162 (2017)
27. Kannadaguli, P., Bhat, V.: A comparison of Bayesian and HMM based approaches in machine learning for emotion detection in native Kannada speaker. In: Proceedings of the IEEMA Engineer infinite conference (eTechNxT), pp. 1–6 (2018)
28. Pulugundla, B., et al.: BUT system for low resource Indian language ASR. In: Interspeech, pp. 3182–3186 (2018)

29. Dua, M., Aggarwal, R.K., Biswas, M.: Discriminative training using noise robust integrated features and refned HMM modeling. J. Intell. Syst. (2018). https://doi.org/10.1515/jisys-2017-0618
30. Samudravijaya, K., Rao, P.V.S., Agrawal, S.S.: Hindi speech database. In: Proceedings of the International Conference on Spoken Language Processing, pp. 456–464 (2002)
31. Fathima, N., Patel, T., Mahima, C., Iyengar, A.: TDNN-based multilingual speech recognition system for low resource Indian languages. In: Proceedings of the Inter-speech, pp. 3197–3201 (2018)
32. Pandey, L., Nathwani, K.: LSTM based attentive fusion of spectral and prosodic information for keyword spotting in Hindi language. In: Interspeech, pp 112–116 (2018)
33. Pal, M., Roy, R., Khan, S., Bepari, M.S., Basu, J.: PannoMulloKathan: voice enabled mobile app for agricultural commodity price dissemination in Bengali language. In: Interspeech, pp. 1491–1492 (2018)
34. Patel, T., Krishna, D.N., Fathima, N., Shah, N., Mahima, C., Kumar, D., Iyengar, A.: Development of large vocabulary speech recognition system with keyword search for Manipuri. In: Proceedings of Inter speech (2018). https://doi.org/10.21437/Interspeech.2018-2133
35. Bhatt, S., Jain, A., Dev, A.: Monophone-based connected word Hindi speech recognition improvement. Sādhanā **46**(2), 1–17 (2021). https://doi.org/10.1007/s12046-021-01614-3
36. Agrawal, S.S., Jain, A., Sinha, S.: Analysis and modeling of acoustic information for automatic dialect classification. Int. J. Speech Technol. **19**(3), 593–609 (2016). https://doi.org/10.1007/s10772-016-9351-7
37. Bhatt, S., Dev, A., Jain, A.: Effects of the dynamic and energy-based feature extraction on Hindi speech recognition. Recent Adv. Comput. Sci. Commun. **14**(5), 1422–1430 (2021)
38. Bhatt, S., Dev, A., Jain, A.: Confusion analysis in phoneme based speech recognition in Hindi. J. Ambient Intell. Humanized Comput. **11**(10), 4213–4238 (2020). https://doi.org/10.1007/s12652-020-01703-x
39. Kumari, R., Dev, A., Kumar, A.: Automatic segmentation of Hindi speech into syllable-like units. Int. J. Adv. Comput. Sci. Appl. **11**(5), 400–406 (2020)
40. Bhatt, S., Jain, A., Dev, A.: Feature extraction techniques with analysis of confusing words for speech recognition in the Hindi language. Wireless Pers. Commun. **118**(4), 3303–3333 (2021). https://doi.org/10.1007/s11277-021-08181-0
41. Bhatt, S., Jain, A., Dev, A.: Continuous speech recognition technologies—a review. In: Singh, M., Rafat, Y. (eds.) Recent Developments in Acoustics. LNME, pp. 85–94. Springer, Singapore (2021). https://doi.org/10.1007/978-981-15-5776-7_8
42. Kumari, R., Dev, A., Kumar, A.: An efficient adaptive artificial neural network-based text to speech synthesizer for Hindi language. Multimedia Tools Appl. **80**(16), 24669–24695 (2021)
43. Sethi, N., Prajapati, D.K.: Text-independent voice authentication system using MFCC features. In: Gupta, D., Khanna, A., Bhattacharyya, S., Hassanien, A.E., Anand, S., Jaiswal, A. (eds.) International Conference on Innovative Computing and Communications. AISC, vol. 1165, pp. 567–577. Springer, Singapore (2021). https://doi.org/10.1007/978-981-15-5113-0_45
44. Sethi, N., Kumar, A., Swami, R.: Automated web development: theme detection and code generation using Mix-NLP. In: ACM International Conference Proceeding Series, p. a45 (2019)
45. Sethi, D., Sethi, N., Gambhir, P., Anand, R.: E-Pandit: automated voice-based system for religious puja's. In: ICRITO 2020 - IEEE 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), pp. 174–181, 9197831 (2020)
46. Sethi, N., Agrawal, P., Madaan, V., Singh, S.K., Kumar, A.: Automated title generation in English language using NLP. Int. J. Control Theor. Appl. **9**(Specialissue11), 5159–5168 (2016)