



Hindi Phoneme Recognition - A Review

Shobha Bhatt¹ (✉), Amita Dev², and Anurag Jain¹

¹ University School of Information and Communication Technology,
GGSIU University, New Delhi, India
anurag@ipu.ac.in

² Indira Gandhi Delhi Technical University for Women, New Delhi, India

Abstract. A review for Hindi phoneme recognition is presented to address Hindi speech recognition. Different issues related to Hindi phonemes such as Hindi speech characteristics, features used in phoneme recognition, and classification method highlighted. Related work was also presented to highlight issues concerned with feature extraction, classification, and distinct features. Earlier reviews mostly addressed speech recognition technologies. This work is an early research study presented for Hindi phoneme recognition. A phoneme-based system is used to overcome the constraint of the requirement of large training samples for word-based models. Phoneme-based systems are widely used for large vocabulary speech recognition, different issues related to consonants and vowels were also included. The comparative analysis is presented for different feature extraction and classification techniques with a recognition score. The research helps by presenting issues related to phoneme recognition.

Keywords: Review · Hindi · Phoneme · Features · Consonants · Vowels

1 Introduction

Speech recognition is the conversion of spoken words into text so that the machine can understand natural voice input to take further action based on the input. Different sub-word models are used in speech recognition to overcome the constraint of large training sample requirements of word-based models. A phoneme-based system is mostly used to overcome the constraint of word-based models due to large training data requirements [1]. Phoneme recognition is one of the fundamental issues in speech recognition. A phoneme is a basic unit for spoken language. The limited number of phonemes can be combined to generate all possible words in a language. Further, due to fewer phonemes, different rules and procedures can be applied for phonemes. Phoneme recognition technologies are applied in speech recognition, speaker identification, language identification, and speech synthesis also[2].

Figure 1 shows the process of speech recognition. The speech recognition systems based on phoneme obtain the highly probable phoneme series for given input speech features [3]. Researchers applied different feature extraction methods and classification methods to recognize the phonemes. Further phoneme recognition can be categorized based on vowel recognition and consonant recognition. The vowels are generated with

an open vocal tract at any point above the glottis. Three articulatory features are high, backness, and roundness. Researchers presented reviews on evolving theories of vowel recognition [4]. Consonants are an important part of any phoneme-based system. Consonants are articulated in the presence of constriction or closure at some point along the vocal tract. The categories of consonants on the basis of place of articulation are bilabial, alveolar, velar, postalveolar, labiodental, and retroflex. Consonants are further categorized on the manner of articulation and voicing. The plosive or stop, fricative, approximant, and closure are grouped under the manner of articulation. The voicing sounds indicate the presence of a voiced region [5]. The recognition of consonants is a challenging task in speech recognition due to their production methods. Researchers made efforts to recognize the consonants using different techniques in the literature.

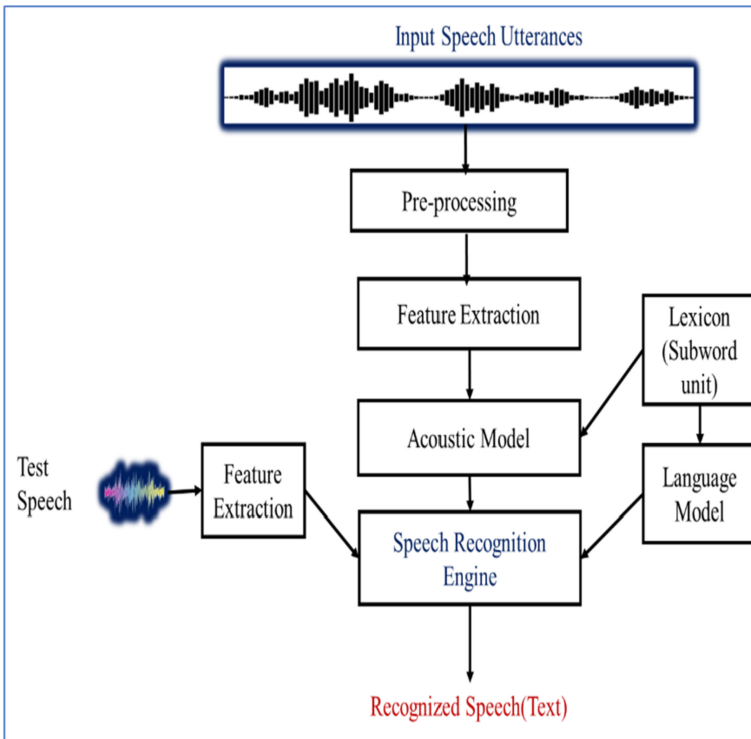


Fig. 1. Speech recognition block diagram

However, phoneme-based speech recognition faces contextual effects. The phonemes are categorized into vowels and consonants. The stop consonants are produced by obstruction of airflow and release. All the periods of stop consonants, such as during which the articulator moves, during which the articulator obstructs the airflow, and when articulators separate to release the air, are important and need to be addressed in the context of speech recognition. Further, recognizing semivowels is also challenging due to acoustically similar characteristics to vowels [6]. Nasalization also causes difficulty

in detection due to antiresonance. Researchers attempted to reduce the contextual effect by using context-dependent triphones [7–9]. Researchers also used longer acoustic units to overcome the contextual effects [10].

Hindi speech recognition systems are discussed in the literature using various feature extraction techniques [2, 11–13]. Different researchers also presented phoneme recognition in Hindi to explore different categories of phonemes such as vowels and consonants. Researchers [2, 14–20] worked for phoneme recognition to find out different issues such as recognition for vowel, consonants, stop consonants, structural analysis, phoneme aspiration detection, retroflex effect, and dental using different feature extraction techniques and classification methods.

Earlier reviews presented for Hindi speech recognition are related to overall speech recognition. The research review in this paper presents different issues related to Hindi phoneme recognition so that researchers can understand underlying issues. The presented review on phoneme recognition will enable researchers to understand different underlying concepts of phoneme recognition related to feature extraction and classification methods to improve Hindi speech recognition. Different Hindi speech characteristics were defined. Further details of feature extraction techniques and classification techniques are provided.

The structure of the paper is detailed as below. The related work is discussed in Sect. 2, while Sect. 3 highlights Hindi language issues. The feature extraction methods are explained in Sect. 4. Section 5 illustrates the classifier, and Sect. 6 is related to results and analysis. Section 7 is the last section with future direction and conclusion.

2 Related Work

The comprehensive review was presented to explore Bangla phonemes for Hidden Markov Model (HMM) and multilayer neural network over a single layer neural network provided [21]. Comparative analysis was also presented. A review of the study of phoneme recognition was presented. Three classifiers HMM, neural network (NN) and vector quantization (VQ) reviewed [22]. Methods described classification and feature extraction techniques for phonemes and isolated word recognition [23]. A survey for phoneme recognition in popular speech corpus was presented for recent deep neural network (DNN) based methods. It was concluded that simple feed-forward DNN provides less phone error rate (PER) compared to other DNN based methods [24]. The study was conducted on important characteristics of pronunciation issues for speech recognition evaluation and comparison [25]. A survey on cross-language for voice onset time (VOT) was addressed for two-stop consonants /d/ and /t/. Finally, VOT was compared for the investigated languages to explore the differences for these stop consonants [26]. A survey was presented for phoneme recognition by the classifier support vector machine (SVM) [27]. A study was presented by using recurrent neural networks (RNNs). It was concluded that RNNs improved speech recognition [28]. A review was presented for landmarks detection such as VOT and burst release to detect stop consonants. It was stated that stop consonants are difficult to recognize due to low energy values, high variabilities, and random behavior [29]. Discriminative phonetic features (DPF) such as the way of phoneme representation were reviewed. It was concluded that Arabic is a Semitic language and needs more research related to DPF. Monophone and hybrid subword units

were used in creating a speech recognition system [30]. The domain-based syntactic structures were also applied to improve speech recognition by reducing the search space during the recognition process. It was observed that maximum word accuracy of 88.54% was achieved with PLP with energy coefficients and a hybrid model. Research findings reveal that substitution errors mostly occurred. Hindi vowel recognition was explored in [2]. The speech recognition framework was implemented using MFCCs with five states of HMM-based modeling. The vowels were subgrouped into the front, back, and middle vowels. The average recognition score of 83.19% was recorded. The results show that accurate prediction of the consonant score was obtained for a broad range of signal-to-noise ratios. The researchers also presented a study to show how vowels and consonants shape the recognition process. It was demonstrated that lexical processing is more strongly connected to consonants in comparison to vowel processing. Acoustically vowels are continuous and long, while consonants are transitory in nature [31]. An extensive study on Hindi phoneme confusion analysis was presented by the researchers [32]. Experiments were carried out using HMM and PLP coefficients on Hindi continuous speech utterances. The results were reported for both consonants and vowels. The vowels attained 70% recognition accuracy. The palatal phonemes achieved the maximum recognition score of 94%.

3 Hindi Language

Hindi alphabets are properly defined [33]. Alphabets in the Hindi language are divided into consonants and vowels. Hindi has about fifty-eight phonemic letters, which include ten vowels, thirty-seven consonants, and an additional five nuktas taken from Farsi/Arabic [34]. Some of the dominant features in the Hindi language are aspiration, gemination, nasalization, and retroflexive [14, 35]. The sounds are voiced and unvoiced in the Hindi language. Hindi vowels are divided into short and long vowels. Table 1 shows the Hindi vowel acoustic classification. Table 2 presents Hindi consonants with IPA symbols [36]. Table 3 presents Hindi semivowels and fricatives with IPA symbols [37].

Table 1. Hindi vowel acoustic classification

Front		Middle		Back	
Hindi vowels	IPA	Hindi Vowels	IPA	Hindi Vowels	IPA
इ	i	अ	ɑ	उ	u
ई	i:	आ	ɑ:	ऊ	u:
ए	e:			ओ	o:
ऐ	ɛ:			औ	ɔ:

Table 2. Hindi consonants

Hindi Consonants	IPA	Label used	Hindi Consonants	IPA	Label
क	K	k	ज	ɲ	ɲj
ख	k ^h	kh	ट	ʈ	t:
ग	G	g	ठ	ʈ ^h	t:h
घ	g ^h	gh	ड	ɖ	d:
ङ	ŋ	ng	ढ	ɖ ^h	d:h
च	tʃ	c	ण	ɳ	n:
छ	tʃ ^h	ch	त	t̪	t
ज	dʒ	j	थ	t̪ ^h	th
झ	dʒ ^h	jh	द	d̪	d
ध	d̪ ^h	dh	ब	b	b
न	n	n	भ	b ^h	bh
प	p	p	म	m	m
फ	p ^h	ph			

Table 3. Hindi semivowels and fricatives

Hindi semivowel	IPA	Label	Hindi fricatives	IPA
य	j	y	श	ʃ
र	r	r	ष	ʂ
ल	l	l	स	s
व	ʋ	v	ह	h

4 Feature Extraction Methods

The vowel recognition was presented using Mel Frequency cepstral coefficients (MFCCs) in [2]. Hindi consonants were classified using EMG-based sub-vocal features in [38] and Linear prediction coefficients [17]. For the recognition of Hindi, phoneme features used are wavelet sub-band based temporal features in [19] and MFCCs in [14]. Researcher work was also presented to develop a phoneme-based system for the Hindi language by using MFCCs, PLPs, and LPCs with their variants [39]. The results show that PLPs and MFCCs performed better than LPCs. The work was also presented using hybrid subword units using PLPs to improve phoneme-based speech recognition [30].

5 Classification Methods

For classification in speech recognition, two models, generative and discriminative, are generally used. The generative models learn from the joint probability distribution of the observed acoustic features and respective speech labels using Bayes rules. In contrast, discriminative training is used to optimize the model parameter [40]. The HMM is simple in design and practical in use for representing variability in speech. However, HMMs are not efficient when modeling nonlinear functions. In contrary to HMM, the ANNs allow discriminative training efficiently. Other research work also presented deep neural networks (DNNs) for improving speech recognition [41]. Several different methods, such as Gaussinization and based on discriminative training, were experimented [42]. Sequence to sequence acoustic modeling proposed for speech recognition [43]. Researchers applied different classification methods for speech recognition based on HMMs. Artificial neural network (ANNs) based classification was applied in [14]. Other works reported Gaussian Mixture Modeling (GMM) for vowel recognition [44]. For recognition of Hindi consonants, vector quantization was used in [17]. The researchers also used context-dependent HMM (CDHMM) for vowel classification [45]. Different matrices have been applied by the researchers to evaluate phoneme recognition. The matrices, such as phoneme error rate (PER), phoneme accuracy, and phoneme correctness, were applied by most of the developers. The phoneme accuracy and phoneme error rate (PER) were used by most of the researchers. The phoneme accuracy and PER are defined is as given below [5, 46, 47]. Hindi phonemes were characterized using time-delay neural networks (TDNNs) [48]. The queries related to Indian railways consisting of 207 Hindi vocabulary words were used in the experiment. Features used in the study were the MFCCs and cepstral mean normalization using the frame. Different TDDNs were trained and tested for Hindi phoneme categorization. Studies also presented to predict consonant recognition and confusion in background noise by using microscopic speech recognition [49].

6 Results and Analysis

The results indicate that researchers worked for the recognition of Hindi phonemes, consonants, and vowels. Most of the works are reported for vowel recognition. Some research findings indicate results for a small group of phonemes. The results were reported for

different categories of phonemes. Researchers used different speech recognition systems. Phoneme recognition was also reported for different environmental conditions. The works were reported for a clean and noisy environment.

MFCCs, LPCs, and wavelet-based methods were applied. Feature extraction techniques based on wavelet sub-band and the combination of wavelet cepstral features with harmonic energy features improved speech recognition. The idea was presented that stops in speech signals are most difficult due to short-duration frequency bursts. Research work also presented different segmentation techniques. The researchers also experimented with subvocal speech recognition based on electromyography signal (EMG). Research also applied Gammatone frequency cepstral coefficients (GFCCs). The following Table 4 shows the list of features used in different research works. The comparative analysis was made based on extracted features, classification methods, phonemes types, and accuracy.

Table 4. Phoneme recognition comparative analysis

S.no	Reference	Features used	Classification method	Recognition	Accuracy
1	[48]	MFCCs	Time delay neural network (TDNN)	Hindi phoneme	Vowel: 99% Unvoiced stops: 87% Voiced stops: 82% Semivowel: 94.7% Fricatives: 96.4% Nasals: 98.1%
2	[2]	MFCCs	HMM	Hindi vowel	Vowel: 83.19%
3	[44]	MFCCs	GMM	Hindi vowel	Microphone recorded speech: 91.4% Telephonic speech: 84.2%
4	[45]	Gammatone frequency cepstral coefficients and formant	CDHMM	Hindi vowel	Speaker dependent: 99.15% Speaker independent: 98.5%
5	[38]	EMG based subvocal features	HMM	Hindi alphabets	/Ka/: 75% /kha/: 78.05% /Ga/: 80.5% /Gha/: 81.3%

(continued)

Table 4. (continued)

S.no	Reference	Features used	Classification method	Recognition	Accuracy
7	[50]	Quantile-based dynamic cepstral normalization MFCC (QCN-MFCC)	HMM	Hindi vowel	Context dependent: MFCC: 89.7% QCN-MFCC: 92.7% Context Independent: MFCC: 87.6% QCN-MFCC: 92.7

Research findings reveal that researchers applied different classification methods. The classification methods used are HMM-based, ANN-based, GMM based, and using vector quantization. Different methods, such as based on backpropagation and time-delay neural network, were applied. Recognition results were presented using accuracy. It was also observed that vowel recognition was mostly explored.

7 Conclusion

A review of Hindi phoneme recognition is presented to understand the issues related to Hindi speech recognition. Different issues related to Hindi phonemes such as Hindi speech characteristics, features used in phoneme recognition, and classification with related work described. The classifiers based on HMM, ANN, GMM, and VQ were experimented. Feature extraction methods improved phoneme recognition. Researchers also worked on subcategories such as vowels and consonants in addition to phonemes. The research work on Hindi speech recognition was also presented using the deep learning method. Researchers also presented studies for phoneme confusion analysis to understand and improve speech recognition. It was also revealed that substitution errors have mostly occurred. Further research may include more studies exploring Hindi phonology and applying hybrid feature extraction methods and classification methods. The outcome of the study consists of that researchers worked mainly on the recognition of the vowels. Further research work may include more studies related to Hindi phonology.

Acknowledgements. The authors would like to acknowledge the Ministry of Electronics & Information Technology (MeitY), Government of India, for providing financial assistance for this research work through “Visvesvaraya Ph.D. Scheme for Electronics & IT”.

References

1. Bhatt, S., Jain, A., Dev, A.: Acoustic modeling in speech recognition: a systematic review. *IJACSA Int. J. Adv. Comput. Sci. Appl.* **11**, 397–412 (2020)

2. Bhatt, S., Dev, A., Jain, A.: Hindi speech vowel recognition using hidden markov model. In: The 6th International Workshop on Spoken Language Technologies for Under-Resourced Languages, pp. 196–199 (2018)
3. Lopes, C., Perdigao, F.: Phoneme recognition on the TIMIT database. *Speech Technol.* (2011). <https://doi.org/10.5772/17600>
4. Strange, W.: Evolving theories of vowel perception. *J. Acoust. Soc. Am.* **85**, 2081–2087 (1989). <https://doi.org/10.1121/1.397860>
5. Vasquez, D., Gruhn, R., Minker, W.: *Hierarchical Neural Network Structures for Phoneme Recognition*. Springer, Berlin (2013). <https://doi.org/10.1007/978-3-642-34425-1>
6. Espy-Wilson, C.Y.: A feature-based semivowel recognition system. *J. Acoust. Soc. Am.* **96**, 65–72 (1994). <https://doi.org/10.1121/1.410375>
7. Bhatt, S., Jain, A., Dev, A.: CICD acoustic modeling based on monophone and triphone for HINDI speech recognition. In: International Conference on Artificial Intelligence and Speech Technology (AIST2019), 14–15th Nov (2019)
8. Mikolov, T., Zweig, G.: Context dependent recurrent neural network language model. In: 2012 IEEE Work Spoken Language *Technology* SLT 2012 – Proceeding, pp. 234–239 (2012). <https://doi.org/10.1109/SLT.2012.6424228>
9. Tüske, Z., Sundermeyer, M., Schlüter, R., Ney, H.: Context-dependent MLPs for LVCSR: TANDEM, hybrid or both? In: 13th Annual Conference of the International Speech Communication Association 2012, INTERSPEECH 2012, vol. 1, pp. 8–21 (2012)
10. Ganapathiraju, A., et al.: Syllable - a promising recognition unit for LVCSR. In: IEEE Workshop on Automatic Speech Recognition and Understanding Proceeding, pp. 207–214 (1997). <https://doi.org/10.1109/asru.1997.659007>
11. Kumar, K., Aggarwal, R.K., Jain, A.: A Hindi speech recognition system for connected words using HTK. *Int. J. Comput. Syst. Eng.* **1**, 25 (2012). <https://doi.org/10.1504/ijcsyse.2012.044740>
12. Sinha, S., Agrawal, S.S., Jain, A.: Continuous density Hidden Markov Model for context dependent Hindi speech recognition. In: International Conference on Advances in Computing, Communications and Informatics, pp. 1953–1958 (2013). <https://doi.org/10.1109/ICACCI.2013.6637481>
13. Pruthi, T., Saksena, S., Das, P.K.: Swaranjali: isolated word recognition for Hindi language using VQ and HMM. *Int. Conf. Multimed. Process. Syst.* **1**, 13–15 (2000)
14. Dev, A.: Effect of retroflex sounds on the recognition of Hindi voiced and unvoiced stops. *AI Soc.* **23**, 603–612 (2009). <https://doi.org/10.1007/s00146-008-0179-9>
15. Sharma, R.P., Khan, I., Farooq, O.: Acoustic study of Hindi unaspirated stop consonants in consonant-vowel (CV) context. *Int. J. Eng. Tech. Res.* **1**, 5–9 (2014)
16. Patil, V.V., Rao, P.: Detection of phonemic aspiration for spoken Hindi pronunciation evaluation. *J. Phon.* **54**, 202–221 (2016). <https://doi.org/10.1016/j.wocn.2015.11.001>
17. Das, P.K., Agrawal, S.S.: Machine recognition of Hindi consonants and distinctive features using vector quantization. *J. Acoust. Soc. Am.* **103**, 2779–2779 (1998). <https://doi.org/10.1121/1.422255>
18. Mishra, A.: Interlaced Derivation for HINDI phoneme- Viseme recognition from continuous speech. *Int. J. Recent Res. Aspects* **4**, 172–176 (2017)
19. Farooq, O., Datta, S., Shrotriya, M.C.: Wavelet sub-band based temporal features for robust hindi phoneme recognition. *Int. J. Wavelets Multiresolut. Inf. Process.* **8**, 847–859 (2010). <https://doi.org/10.1142/S0219691310003845>
20. Khan, M., Jahan, M.: Classification of myoelectric signal for sub-vocal Hindi phoneme speech recognition. *J. Intell. Fuzzy Syst.* **35**, 5585–5592 (2018). <https://doi.org/10.3233/JIFS-161067>
21. Tasnim Swarna, S., Ehsan, S., Islam, S., Jannat, M.E.: A comprehensive survey on bengali phoneme recognition. In: Proceedings of the International Conference on Engineering Research, Innovation and Education 2017 ICERIE 2017, pp. 1–7 (2017)

22. Kshirsagar, A., Dighe, A., Nagar, K., Patidar, M.: Comparative study of phoneme recognition techniques. In: *Proceeding of 2012 3rd International Conference on Computer and Communication Technologies ICCCT 2012*, pp. 98–103 (2012). <https://doi.org/10.1109/ICCCT.2012.228>
23. Yusnita, M.A., Paulraj, M.P., Yaacob, S., Abu Bakar, S., Saidatul, A., Abdullah, A.N.: Phoneme-based or isolated-word modeling speech recognition system? An overview. In: *Proceedings - 2011 IEEE 7th International Colloquium on Signal Processing and Its Applications, CSPA 2011*, pp. 304–309 (2011). <https://doi.org/10.1109/CSPA.2011.5759892>
24. Michálek, J., Vaněk, J.: A survey of recent DNN architectures on the TIMIT phone recognition task. In: Sojka, P., Horák, A., Kopeček, I., Pala, K. (eds.) *Text, Speech, and Dialogue. TSD 2018. Lecture Notes in Computer Science*, vol 11107. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00794-2_47
25. Strik, H., Cucchiaroni, C.: Modeling pronunciation variation for ASR: a survey of the literature. *Speech Commun.* **29**, 225–246 (1999). [https://doi.org/10.1016/S0167-6393\(99\)00038-2](https://doi.org/10.1016/S0167-6393(99)00038-2)
26. AlDahri, S.S., Alotaibi, Y.A.: A crosslanguage survey of VOT values for stops (/d/, /t/). In: *Proceeding - 2010 IEEE International Conference on Intelligent Computing and Intelligent Systems, ICIS 2010*, vol. 3, pp. 334–338 (2010). <https://doi.org/10.1109/ICICISYS.2010.5658744>
27. Fathima Nazarath, P.A.: Survey on phoneme recognition using support vector machine. In: *National Conference on Emerging Research Trend in Electrical and Electronics Engineering (ERTE 19)*, pp. 187–192 (2019)
28. Koizumi, T., Mori, M., Taniguchi, S., Maruya, M.: Recurrent neural networks for phoneme recognition. In: *International Conference Spoken language processing, ICSLP, Proceeding*, vol. 1, pp. 326–329 (1996). <https://doi.org/10.1109/icslp.1996.607119>
29. Nirmala, S.R., Upashana, G.: Advances in computational research a review on landmark detection methodologies of stop consonants. *Adv. Comput. Res.* **8**, 316–320 (2017)
30. Bhatt, S., Jain, A., Dev, A.: Monophone-based connected word Hindi speech recognition improvement. *Sādhanā* **46**, 1–17 (2021). <https://doi.org/10.1007/S12046-021-01614-3>
31. Nazzi, T., Cutler, A.: How consonants and vowels shape spoken-language recognition. *Annu. Rev. Linguistics* **5**, 25–47 (2018). <https://doi.org/10.1146/annurev-linguistics>
32. Bhatt, S., Dev, A., Jain, A.: Confusion analysis in phoneme based speech recognition in Hindi. *J. Ambient Intell. Humaniz. Comput.* **11**, 4213–4238 (2020). <https://doi.org/10.1007/s12652-020-01703-x>
33. Bansal, P., Dev, A., Jain, S.B.: Optimum HMM combined with vector quantization for Hindi speech recognition. *IETE J. Res.* **54**, 239–243 (2008). <https://doi.org/10.4103/0377-2063.44216>
34. Aarti, B., Kopparapu, S.K.: Spoken Indian language identification: a review of features and databases. *Sadhana - Acad. Proc. Eng. Sci.* **43**, 1–14 (2018). <https://doi.org/10.1007/s12046-018-0841-y>
35. Malviya, S., Mishra, R., Tiwary, U.S.: Structural analysis of Hindi phonetics and a method for extraction of phonetically rich sentences from a very large Hindi text corpus. In: *Conference of the Oriental Chapter of International Committee for Coordination and Standardization of Speech Databases and Assessment Techniques, O-COCOSDA 2016*. pp. 188–193 (2017). <https://doi.org/10.1109/ICSODA.2016.7919009>
36. Sadhukhan, T., Bansal, S., Kumar, A.: Automatic identification of spoken language. *IOSR J. Comput. Eng.* **19**, 84–89 (2017). <https://doi.org/10.9790/0661-1902058489>
37. Kachru, Y.: *Hindi*. John Benjamins Publishing, London (2006)
38. Khan, M., Jahan, M.: Sub-vocal speech pattern recognition of Hindi alphabet with surface electromyography signal. *Perspect. Sci.* **8**, 558–560 (2016). <https://doi.org/10.1016/j.pisc.2016.06.019>

39. Bhatt, S., Jain, A., Dev, A.: Feature extraction techniques with analysis of confusing words for speech recognition in the Hindi language. *Wirel. Pers. Commun.* **118**, 3303–3333 (2021). <https://doi.org/10.1007/S11277-021-08181-0>
40. Gales, M.J.F., Watanabe, S., Fosler-Lussier, E.: Structured discriminative models for speech recognition: an overview. *IEEE Sig. Process. Mag.* **29**, 70–81 (2012). <https://doi.org/10.1109/MSP.2012.2207140>
41. Wason, R.: Deep learning: evolution and expansion. *Cogn. Syst. Res.* **52**, 701–708 (2018). <https://doi.org/10.1016/j.cogsys.2018.08.023>
42. Liu, X., Gales, M.J.F., Sim, K.C., Yu, K.: Investigation of acoustic modeling techniques for LVCSR systems. In: *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing I* (2005). <https://doi.org/10.1109/ICASSP.2005.1415247>
43. Zhang, J.X., Ling, Z.H., Liu, L.J., Jiang, Y., Dai, L.R.: Sequence-to-sequence acoustic modeling for voice conversion. *IEEE/ACM Trans. Audio Speech Lang. Process.* **27**, 631–644 (2019). <https://doi.org/10.1109/TASLP.2019.2892235>
44. Koolagudi, S.G., Thakur, S.N., Barthwal, A., Singh, M.K., Rawat, R., Sreenivasa Rao, K.: Vowel recognition from telephonic speech using MFCCs and Gaussian mixture models. In: *Communications in Computer and Information Science*, pp. 170–177 (2012). https://doi.org/10.1007/978-3-642-32112-2_21
45. Biswas, A., Sahu, P.K., Bhowmick, A., Chandra, M.: Hindi vowel classification using GFCC and formant analysis in sensor mismatch condition. *WSEAS Trans. Syst.* **13**, 130–143 (2014)
46. Moses, D.A., Mesgarani, N., Leonard, M.K., Chang, E.F.: Neural speech recognition: continuous phoneme decoding using spatiotemporal representations of human cortical activity. *J. Neural Eng.* **13**, 056004 (2016). <https://doi.org/10.1088/1741-2560/13/5/056004>
47. Gales, M., Young, S.: The application of hidden Markov models in speech recognition. *Found. Trends Sig. Process.* **1**, 195–304 (2007). <https://doi.org/10.1561/20000000004>
48. Dev, A., Agrawal, S.S., Choudhury, D.R.: Categorization of Hindi phonemes by neural networks. *AI Soc.* **17**, 375–382 (2003). <https://doi.org/10.1007/s00146-003-0263-0>
49. Zaar, J., Dau, T.: Predicting consonant recognition and confusions in normal-hearing listeners. *J. Acoust. Soc. Am.* **141**, 1051–1064 (2017). <https://doi.org/10.1121/1.4976054>
50. Mishra, S., Bhowmick, A., Shrotriya, M.C.: Hindi vowel classification using QCN-MFCC features. *Perspect. Sci.* **8**, 28–31 (2016). <https://doi.org/10.1016/j.pisc.2016.01.010>