Acoustic Modeling in Speech Recognition: A Systematic Review

Shobha Bhatt[1], Anurag Jain[2]
University School of Information and Communication Technology, Guru Gobind Singh Indraprastha University(GGSIPU), New Delhi, India

Amita Dev[3]
Indira Gandhi Delhi Technical University for women
Dept. name of organization
New Delhi, India

***Abstract*—The paper presents a systematic review of acoustic modeling (AM) techniques in speech recognition(SR). Acoustic modeling establishes a relationship between acoustic information and language construct in SR. Over the past decades, researchers presented studies addressing specific concerns in AM. However, all previous research works lack a systematic and comprehensive review of acoustic modeling issues. A systematic review is introduced to understand the acoustic modeling issues in speech recognition. This paper provides an extensive and comprehensive inspection of various researches that have been performed since 1984. The extensive investigation and analysis into AM was performed by getting the relevant data from 73 research works chose after the screening process between the years from 1984 to 2020. The systematic review process was divided into different parts to investigate acoustic modeling issues. Main issues in acoustic modeling such as feature extraction techniques, acoustic modeling units, speech corpora, classification methods, different tools used, language issues applied, and evaluation parameters were investigated. This study helps the reader to understand various acoustic modeling issues with comprehensive details. The research outcomes presented in this study depict research trends and shed light on new research topics in AM. The result of this review can be used to build a better speech recognition system by choosing a suitable acoustic modeling construct in SR.**

***Keywords*—Acoustic modeling; speech recognition; systematic review; acoustic unit; MFCC; classification**

I. Introduction

This Speech Recognition(SR) is intended to convert spoken term into text. Nowadays, with an increasing number of devices, people are using a speech recognition system such as Siri with iPhone, Alexa from Amazon, and Cortana for windows. Speech recognition systems are becoming popular due to different commercial and personal purposes [1]. As speech recognition is influencing every field of life, so it has been a concern of researchers as humans always wanted to talk to machines. Speech recognition understanding systems have helped human beings in different ways. In recent years, researchers also started experimenting to learn human activities from audiovisual inputs using neural networks even. Speech recognition systems are applied in speech-enabled devices, medical, machine translation systems, home automation systems, and the education system [2].

Acoustic Modeling is an initial and essential process in speech recognition. The acoustic model establishes the relation between acoustic information and linguistic unit. Most of the calculations are performed in acoustic modeling due to feature extraction and statistical representation, so it primarily affects the recognition process. Statistical representations are prepared from extracted features. The distribution of extracted features with particular sound is modeled in AM to establish the link between extracted features and structures of the linguistic unit. Various feature extraction techniques, such as based on human perception and working of voice production mechanisms, have been reported[3]–[5]. Features were extracted for AM in speaker-independent mode recognition as these systems impose difficulties in speech recognition [6]–[9].

For developing acoustic models, the selection of classification methods is also an important step. Many research works have been reported for acoustic modeling based on different classification techniques[10]. The research work reported using different classification methods such as based on hidden Markov model(HMM), discriminative training for optimization of the model parameter, artificial neural networks(ANNs), deep neural networks(DNNs), and sequence to sequence acoustic modeling.

Further, AM is also linked to many concepts. It requires an understanding of the acoustic-phonetic knowledge, microphone and environment variability issues, gender, and dialectal differences. Further, for determining the connection between linguistic units and acoustic observation, rigorous training is required [11]. AM is also directly linked to pronunciation modeling, variability modeling related to speaker, environment, and contexts also [12]. Acoustic models using subword units also experimented for recognition enhancements [13]. The subword modeling units, such as phone diaphones, syllables, and context-dependent phones, were used [5]. It was also reported that phoneme based models are used to overcome a huge quantity of data for creating trained models. Different models, such as context-dependent, also experimented. Triphone based context-dependent models were used to reduce contextual effects [14], [15]. Researchers also addressed acoustic modeling for a multilingual SR system by using clustering with decision trees taking advantage of the data in the languages other than target language [16]. Further, different language modeling techniques are used in speech recognition. N-gram models are widely used that can model word prediction based on probability [17].

AM in SR also faces different challenges. The task of acoustic modeling is complicated, as well as exciting [18]. The

design of adequate modeling has been a constant effort from the starting of Automatic Speech Recognition (ASR) [19]. The problem of data scarcity has always been a concern for the researchers. Researchers and different groups have developed different speech corpora as per the requirements. However, still, researchers are facing the lack of speech corpus in the public domain, especially for low resource languages for the realization of recognition frame works [20]. Researchers developed acoustic modeling methods using deep neural networks (DNNs) for zero resources language for unsupervised SR [21]. The selection of feature extraction in mismatch and noisy condition makes acoustic modeling a challenging task. Researchers experimented with different robust feature extraction techniques with further processing in acoustic modeling for different environmental conditions. Additional acoustic modeling task is complicated due to contextual variability, pronunciation variability, and speaker variability. Researchers attempted to improve speech recognition by using different acoustic units, robust feature extraction, and different classification methods [13], [22]–[25].

During the past decades, researchers have presented reviews on different acoustic modeling techniques in SR. However, most of the researchers focused only on some specific issues in acoustic modeling and did not cover all the key issues in acoustic modeling. Very less paper has been reported, which shows a complete and systematic review of acoustic modeling in speech recognition. There is a need for systematic analysis of the earlier presented research works to elaborate basic and advanced concepts in acoustic modeling. This work intended to show a systematic literature review(SLR) to meet this gap and to provide a thorough review of AM issues for both novice users and specialists in the field of SR.We presented a comprehensive study in this field. Specifically, we emphasized the key issues related to feature extraction techniques, classification techniques, acoustic modeling units, speech corpora, language issues, different tools, and evaluation parameters for investigation. The research methodology used in this study has been adopted from [26]–[30]. The systematic review process was divided into requirement analysis for systematic review, the setting of research questions, formulation of searching criteria for research papers, the process of paper selection and rejection, setting of assessment measures for the collection of the papers in a systematic review, extraction of relevant information as per the research questions, and finally reporting the results with analysis and discussion. The research investigation focused on acoustic modeling issues in speech recognition by a comprehensive study of 73 research papers extracted from the research works between 1984 to 2020. A total of 127 papers were selected for the complete survey after the initial screening of 250 papers, out of 127 papers, 73 papers were selected for the systematic review process. Different research questions were framed to address acoustic modeling issues, and answers were provided by extracting relevant information from the research papers. With this review, we provide the speech research community with the understanding to decide among acoustic modeling methods as per the requirement. To better understand the AM concepts, we have also described the basic concepts in speech recognition and acoustic modeling. Research findings show different research trends and highlight new research areas. The advantages and disadvantages of various issues are also provided as a guide to interested new and experienced researchers. We have attempted to address all possible aspects of acoustic modeling. Throughout this paper, a constant effort has been made to address issues in a comprehensive way to fill the research gaps. The paper contributed by exploring the following facts.

- Different feature extraction techniques for AM explored.
- Various classification techniques for AM identified.
- The need and different characteristics of speech corpora revealed.
- Different software and tools explored.
- Acoustic modeling units investigated.
- Various language issues used in speech recognition for AM identified.
- The types of publication (Journal, conference, workshops, lecture notes, thesis) identified.
- The specific names of the journal or conference that published the paper.
- Different evaluation criteria defined.

The paper is structured as follows. Section II depicts related work. The speech recognition process and acoustic modeling is elaborated in section III. Section IV clarifies the methodology for the systematic review. Section V is about result and analysis. Section VI depicts discussion. Last section finishes up with conclusion and future direction.

II. Related Works

Reviews on various issues, including acoustic modeling, have been presented for the SR framework. Acoustic modelings with the acoustic-phonetic methodology and pattern recognition methods were addressed in[31]. Researchers discussed several factors to enhance SR. The factors include the usage of HMM modeling, the use of sub-word models, and corrective training. The focus of the paper was on the use of subword models with or without context dependency-based AM modeling. The researchers experimented with several methods to create acoustic models to characterize phone like units. The context-dependent modeling improved the recognition results.

Researchers presented a review of HMM-based speech recognition [32]. The study covered HMM architecture, different techniques, and related issues. The developers included different parameters such as the selection of optimal states, number of gaussian mixture models, context-dependent and triphone based modeling, feature vector, selection of speech databases, and speech-language model. The widely known HMM-based tool kit HTK was explained. It was concluded that HMM-based speech recognition technology was widely used and accepted by the researchers for the decades on a large scale.

Research work was presented for noisy conditions using a taxonomy-based approach [33]. The authors used different key attributes to offer insight into noise-robust methods in SR. The survey addressed the techniques which were successful over the years and had the future for further research. Further techniques were evaluated using five different criteria. The first measure was based on feature space versus the model domain to analyze the mismatch in training and testing conditions. The second criterion was based on compansion using formal information about acoustic distortion. The third criteria were regarding compensation with implicit versus explicit distortion modeling. The fourth criterion was based on uncertainty versus deterministic processing. The fifth criterion was based on the joint model, preparing versus disjoint preparation.

An overview of different modeling techniques such as hidden Markov models (HMMs), Deep neural networks (DNNs) and convolution neural networks (CNNs) was covered [34]. The advanced features of CNN architecture were also discussed. The advantages of using CNNs such as normalization of speaker variances by using local filters in the convolution layer were elaborated. It was concluded that in this decade, the researchers are focussing on DNNs and CNNs for acoustic modeling to overcome the challenges in the SR systems.

The study focused on the comparison of feature extraction, classification, and language models used in SR [35]. The paper was started with a description of the basic SR framework and with its key elements. Different popular feature extraction techniques such as Mel Frequency Cepstral Coefficients (MFCCs), Perceptual Linear Prediction (PLP) Cepstral Coefficients, Relative Spectral Perceptual Linear Prediction Coefficients (RASTA-PLP), Linear Prediction Cepstral (LPC), Discrete wavelet transform (DWT) and transformation techniques were applied. It was stated that MFCCs is widely used and renowned features. The classification method, such as HMMs, the ANNs, and SVMs described for the ASR system. It was stated that the hybrid approach of combining HMMs with other models is being experimented with by researchers. Findings also indicate that SVM based speech recognition systems are also being adopted due to their better performance than ANNs. Finally, it was stated that spoken language also affects the speech recognition process. The comparative study of different issues was also presented to understand the topic better.

The review paper on machine learning (ML) in ASR presented [36]. The ML techniques in speech recognition discussed and provided insights into the ML paradigm in the SR process. Different machine learning approaches GMM-HMM, ANN, support vector machines (SVM), and Deep learning techniques described with their characteristics. Fundamentals concepts of neural networks also explained. It was concluded that ML techniques are widely being experimented in speech recognition, and recent advancements in deep learning work like Connectionist temporal classification (CTC) based acoustic modeling is an exciting path towards continuous speech recognition for large vocabularies.

A review paper was presented to address acoustic modeling issues and refinements [37]. The first constructs and functioning of HMM and its constraints reviewed. Further advancements and improvements to conventional HMM were also explored. The current challenges and performance issues to speech recognition systems also investigated.

A survey of speech recognition using Deep Neural Networks(DNNs) was presented [30]. Research findings include the data related to different databases used, various feature extraction techniques, and modeling techniques. It was stated that for speech corpus, both public and private databases were used. The speech recognition systems were applied to different environments, such as noise, neutral, and emotional. Researchers discussed the use of different classification methods such as Deep neural networks (DNNs), Deep Belief Networks (DBN), Convolution neural networks (CNNs), Recurrent Neural Networks (RNNs), Deep Max out networks (DMN), Deep Convex Network (DCN), Deep stacking network (DSN), Deep Tensor network (DSN) and autoencoder in speech recognition systems.

The brain spiking neural networks (SNNs) were applied to explore large vocabulary speech recognition [38]. These networks are inspired by the brain working and have low computation cost. The work is the progress towards rapid and energy-efficient SR. The ASR can be developed using PyTorch, and it can be easily associated with the PyTorch-Kaldi speech recognition tool kit. The results show that the system provided better accuracy than their ANN counterparts. The time-delay neural network-based acoustic modeling presented for Hindi speech recognition [39]. It was indicated that TDNN showed improvement over GMM-HMM systems.

The presented work differs from the above-mentioned reviews, as we have given a detailed and thorough examination of the acoustic modeling and its related issues in speech recognition systems. The paper first provided an overview of speech recognition and AM. This study provided the reader with the appropriate background to fully understand the topic presented. The systematic review was carried out by using papers from 1984 to 2020. We have introduced a systematic review by including the research works from the beginning, middle, and recent years to understand the flow of acoustic modeling research in speech recognition.

III. Speech Recognition Process and Acoustic Modeling

A generalized speech recognition system includes preprocessing, feature extraction, acoustic modeling, and language modeling units with a recognition engine. Fig. 1 illustrates the SR framework with two phases. The complete recognition process was divided into two components acoustic analysis and acoustic/linguistic decoder. The preprocessing block consists of pre-emphasis to increase the magnitude of higher frequencies to flatten the magnitude spectrum and windowing of speech signals [40]. By applying to the window, a small segment of the speech signal, which is considered as stationary for speech processing analysis, is extracted [41]. The output of feature extraction block is feature vectors which are further used in acoustic modeling of speech utterances. The acoustic model is prepared from the speech database and

linguistic construct. The language model block contains all the programs related to the language modeling issues required for speech recognition. During the recognition phase of speech, word sequences probability is estimated by the language model (LM). Further, language models are used in speech recognition to make a decision regarding acoustically confused spoken utterances by incorporating syntactical and semantic constraints of the spoken language [42], [43]. It also restricts the search space of the recognition engine [44]. The speech recognition process finds the best sequence of words based on the acoustic model, language model, and recognition engine.

The development and design of speech corpus is an essential step towards acoustic modelling [43]. As nowadays, speech recognition systems are being developed for various needs, so the design and development of speech databases play a crucial role in acoustic modeling. The phonetic information is extracted for acoustic modeling from speech corpus. Speech corpora are also used to train and test recognition systems. Further, it is also an important decision to select the acoustic unit in acoustic modeling. Researchers used word-level acoustic modeling; however, there is always a problem of data scarcity in word-level acoustic modeling. The sub-word models are applied to overcome the requirement of a large number of word instances in training for word-based models. The subword models, such as based on phoneme, syllable, and triphones, are commonly used [45].

The phoneme based models are used to overcome the more training data requirement due to word-based models, especially in continuous speech recognition designed for enormous vocabulary size. The phoneme based system suffers from contextual effects. The contextual effects ere reduced by using triphone based systems that consider the left and right contexts of the phonemes. Triphone based systems suffer from data scarcity. The syllable based system is used to cover a larger acoustic unit [46]–[48] to reduce the contextual effects due to phoneme based system. Researchers also attempted to use universal phone sets for multilingual speech recognition and under resource languages [49], [50].

During feature extraction, the insignificant information is removed from the speech signal. Various methods based on speech perception and production have been applied, such as LPC, MFCCs, and PLP [51], [52]. Researchers have also worked to find features for different environments and speaker-independent systems. Different noise-robust feature extraction techniques applied. Acoustic models also generated from extracting features from spectrogram images using convolutional neural networks (CNNs) [53].

In acoustic modeling, different classification methods are used. Automatic speech recognition classification methodologies can be categorized based on acoustic-phonetic knowledge, concepts on pattern recognition, and artificial intelligence(AI) [54], [55]. Widely used techniques are based on Hidden Markov Models (HMMs) and artificial neural networks(ANNs). Discriminative training is also used, which includes both feature extraction and classification in order to provide the minimization of classification errors. It ensures that the classifier will itself map an input space to more suitable for its proper classification [56]–[58].

The recent works have been reported using deep learning-based acoustic models. The researchers generated acoustic word models using contextual information for long conversational speech using a joint CTC/attention-based approach [59]. Speech recognition also improved by using Long Short Term Neural Networks(LSTM) based on language modelling [43]. Researchers investigated DNN based models obtained up to 30% relative error reduction over best discriminatively trained GMMs. The performance of the DNN based system is also influenced by feature vectors used [60].

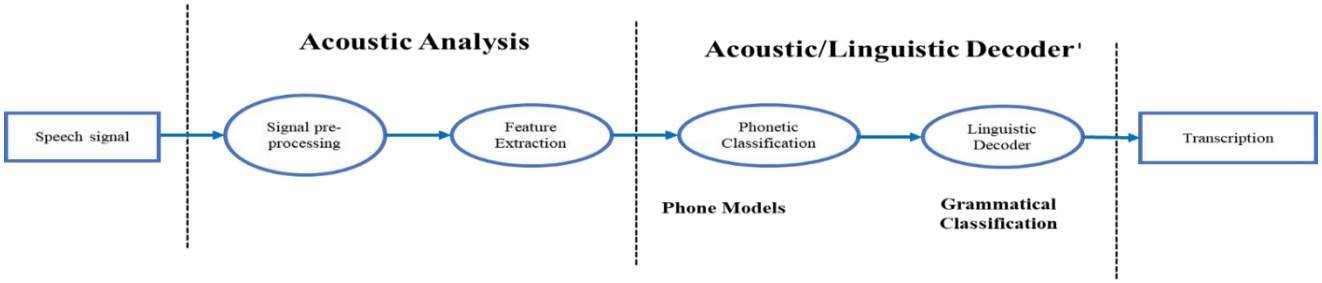

Fig. 1. Speech Recognition Process with Two Phases Acoustic Analysis and Acoustic/Linguistic Decoder[44].

IV. Methodology for a Systematic Review

The systematic review conducted in this paper is based on studies [27]–[29], [61]. They have divided the investigation into the planning phase, executing phase, and finally reporting phase. We have grouped the systematic review process into eight steps. Fig. 2 shows the methodology used to perform a systematic review.

The review process started with a requirement analysis of the systematic study in acoustic modeling. The second phase included identifying and formulating research questions as per our defined goals and gaps based on earlier surveys. The strategy to search the papers from different resources was decided in the third phase. The fourth phase is about inclusion and exclusion criteria for the determination of the research papers. The evaluation criteria for the final selection of the papers for the systematic review were prepared in the fifth phase. The sixth phase was regarding collecting the data from extracted papers. The results were reported in the seventh phase. The last phase presented evaluation and analysis. The following subsections demonstrate the review protocols used in this study in detail.

A. Formulation of Research Questions

To meet our goal of the study, different research questions were framed to conduct a systematic review. Various issues discussed are related to study papers utilized, types SR system used, language applied, language issue covered, speech corpora used, software and tools used, acoustic units experimented, extraction features utilized, classification methods used, and performance metrics applied. Table I lists the research questions. A total of ten research questions were formulated to reveal different aspects of AM in speech recognition.

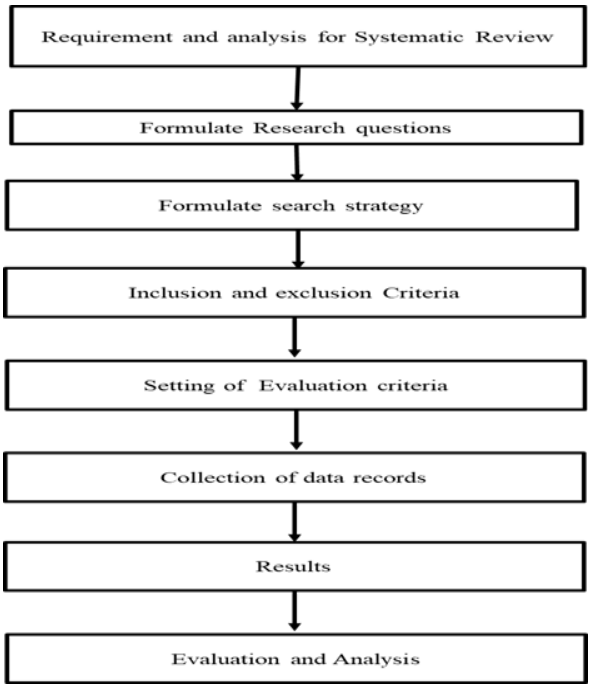

Fig. 2. Methodology for Systematic Review of Acoustic Modelling in Speech Recognition.

TABLE I. Research Question used in a Systematic Review in AM

Slno.	Research question
RQ1	Which type of research papers used in the study?
RQ2	Which type of speech recognition systems identified?
RQ3	What are the languages found in the research investigation?
RQ4	Which are the various language issues used in speech recognition for acoustic modeling?
RQ5	What are the different databases used in the study?
RQ6	What are the different software and tools found in the inspection of the works?
RQ7	Which are the different acoustic modeling units used in the study?
RQ8	What are the different feature extraction techniques used in acoustic modeling?
RQ9	Which different classification techniques are used in speech recognition?
RQ10	What are different performance measurements in the speech recognition system

B. Search Strategy

For searching the research papers, all the key terms related to research questions were used. Further exploration was also done based on specific journals related to speech processing. Different connectors, such as 'OR' and 'AND' were used. Various resources such as Google search, Google scholar, IEEE explore, Springer, Taylor, and Francis, research within specified journals such as Speech Communication, Science Direct, university repositories for thesis, lecture notes, and books were searched.

C. Study Selection

Initially, we extracted a total of two hundred fifty papers. All replica papers and the same principles papers were eliminated. After this step, inclusion and exclusion criteria were applied. The papers were excluded, which contained speaker recognition and emotion recognition. The papers which were related to speech processing but do not contain acoustic modeling issues were also not selected. Papers related to acoustic modeling issues in speech recognition were selected, and papers for acoustic modeling for different acoustic units were also included. Then finally, a total of 127 papers were decided for the study.

D. Quality Assessment Criteria

The research papers for systematic review were chosen at last subsequent to applying quality assessment criteria on the explored papers got after inclusion and exclusion parameters, as discussed in the study selection section. The quality assessment criteria were based on 21 questions. Table II lists the quality questions used for the evaluation of a systematic research review. The following quality assessment rules were applied for the selection of the papers.

Rule1: If the answer meets the full requirement, it is awarded 1.

Rule2: If the question is not answered, it is awarded 0.

Rule3: If the answer is satisfactory, it is awarded 0.5.

Rule4: If the answer is above average, it is awarded 0.75.

Rule5: If the answer is below average, then it is awarded 0.25.

TABLE II. QUALITY ASSESSMENT QUESTIONS FOR THE FINAL SELECTION OF THE PAPERS IN THE STUDY

Sl.no.	Questions for quality assessment
QA1.	Are the objectives of the study clearly stated?
QA2	Are challenges and gaps mentioned?
QA3	Does the need for research clearly stated?
QA4	Does the research include an incremental contribution to the researchers?
QA5	Is the speech recognition system process mentioned?
QA6	Is the experimental setup mentioned?
QA7	Is the speech corpus is appropriate and clearly defined?
QA8	Are the feature extraction techniques defined?
QA9	Are acoustic units clearly defined?
QA10	Are the classification techniques specified?
QA11	Are the tools and software for developing speech recognition stated clearly?
QA12	Is any language modeling method applied?
QA13	What are the different search strategies applied?
QA14	Is any research finding is available for improving speech recognition?
QA15	What are the different metrics used for the performance emeasurement of the developed SR framework?
QA16	Are all the results specified shown?
QA17	Is there a separate analysis section?
QA18	Do experiments support all research findings?
QA19	Is there any comparative analysis conducted?
QA20	Does the literature review is relevant and addresses the research question?
QA21	Was the paper useful?

Then for every paper, the summation of marks is added for all 21 questions. We have included all the papers which got a score of 13 or above marks. Other papers were excluded from the study. Finally, we have included only 73 research papers.

V. RESULTS AND ANALYSIS

The systematic review process aimed at the investigation of AM issues in speech recognition. Research questions were framed, and relevant data were extracted to get the solutions for these questions from RQ1 to RQ10. The outcome of the study covers all the important concerning areas for acoustic modeling. The following sections describe the research outcomes with analysis.

A. RQ1 Aimed to Find the Various Type of Research Papers used in the Study

The papers were selected after applying quality assessment criteria. Different search directories were used, and finally, 73 research papers were included for the systematic review from the year 1984 to 2020. Fig. 3 shows the year-wise distribution of the papers. Table III shows the overall distribution of the papers among Journal/Conferences/Workshops/Lecture Notes /Thesis. It was observed that conferences provided the highest 45% of the papers, while journal papers show 41% participation. The papers from the workshop show 8% participation, while lecture series and thesis both show only 3% participation individually. Further analysis was made to investigate the Journal/Conferences/Workshops/Lecture Notes /Thesis papers independently.

Table IV indicates the list of the Journals and their overall percentage. It was seen that the journals “IEEE Transaction Speech Audio Processing” and “International Journal of Speech and Technology” given the highest number of papers in the study.

Table V indicates the list of conferences and their overall percentage. It was observed that the conferences “ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing,” “Proceedings of the European Conference on Speech Communication and Technology,” and “ INTERSPEECH” provided the highest number of the papers.

Table VI shows the name of the workshops and their overall percentage. Table VII indicates the list of the lecture series and their overall percentage. Table VIII indicates the name of the universities with the published thesis and their overall percentage. It was additionally discovered that very fewer papers reported from the workshops, lecture series, and thesis.

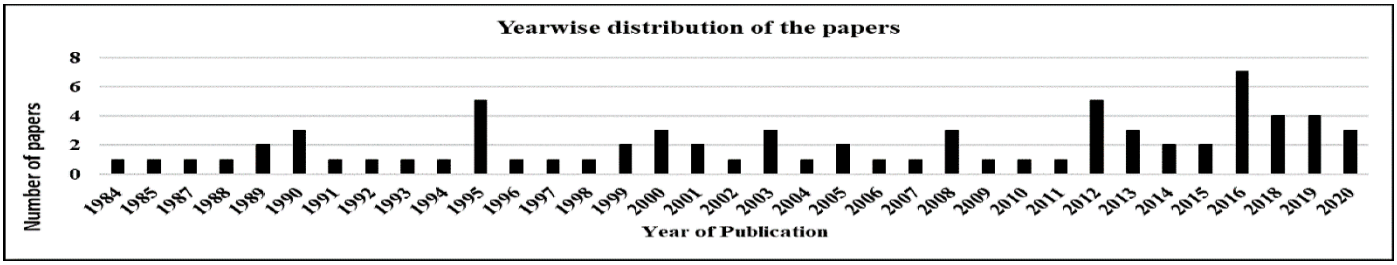

Fig. 3. Yearwise Distribution of the Papers Selected for Systematic Review in Acoustic Modeling for Speech Recognition between the Years 1984-2020.

TABLE III. DISTRIBUTION OF PAPERS AMONG JOURNAL/CONFERENCES/WORKSHOPS/LECTURE NOTES /THESIS USED IN THE STUDY WITH REFERENCE NUMBERS

Type of the papers	Number of papers	Percentage of type of paper(%)	Paper references
Journals	30	41.1	[62], [63], [64], [65], [66], [67], [68], [69], [70], [44], [71], [72], [73], [74], [75], [76], [77], [78], [79], [80], [47], [81], [82], [83], [84], [85], [53], [86], [87], [5]
Conferences	33	45.2	[88], [63], [89], [90], [91], [92], [93], [94], [95], [96], [97], [98], [99], [100], [22], [101], [102], [103], [104], [105], [106], [107], [108], [109], [110], [111], [112], [113], [114], [115], [116], [117], [118]
Workshops	6	8.22	[119] ,[120], [121], [46], [122], [123]
Lecture Notes	2	2.74	[124], [125]
Thesis	2	2.74	[126], [127]
	73		

TABLE IV. NAME OF THE JOURNALS, NUMBER OF THE PAPERS, PERCENTAGE OF THE PAPERS USED IN THE SYSTEMATIC REVIEW FOR AM IN SPEECH RECOGNITION

Sl.no	Name of the Journal	Total papers	Journal paper(%))
1.	IEEE Trans Speech Audio Processing	4	5.48
2.	IETE Journal of Research	1	1.37
3.	International Journal of Speech Technology	4	5.48
4.	Journal of Brazilian Computer Society	1	1.37
5.	Journal of Shanghai University	1	1.37
6.	Procedia Engineering	1	1.37
7.	Advanced Materials Research	1	1.37
8.	South African Computer Journal	1	1.37
9.	AI Society	1	1.37
10.	Recent Advances in Computer Science and Communication	1	1.37
11.	Speech Communication	2	2.74
12.	Advances in Intelligent Systems and Computing. Springer	1	1.37
13.	AT&T Bell Lab Technical Journal	1	1.37
14.	European Student Journal of Language and Speech	1	1.37
15.	Eurasip Journal of Audio, Speech, Music Processing	1	1.37
16.	IEEE ASSP Magzine	1	1.37
17.	International Journal of Computational Systems and Engineering	1	1.37
18.	Journal of Ambient Intelligence and Humanized Computing	1	1.37
19.	Neurocomputing. Springer Berlin Heidelberg	1	1.37
20.	WSEAS Transaction Signal Processing	1	1.37
21.	IEEE Access	1	1.37
22.	International Arab Journal of Information and Technology	1	1.37
23.	IOSR Journal of VLSI and Signal Processing	1	1.37

TABLE V. NAME OF THE CONFERENCES, NUMBER OF THE PAPERS, PERCENTAGE OF THE PAPERS USED IN THE SYSTEMATIC REVIEW FOR AM IN SPEECH RECOGNITION

Sl.no.	Name of the Conferences	Total papers	Conference papers(%)
1	International Conference on Trends in Automation, Communication, and Computing Technologies (I-TACT). Institute of Electrical and Electronics Engineers Inc.	1	1.37
2	International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE	8	11
4	European Conference on Speech Communication and Technology	8	10.96
5	International Conference on Emerging Trends in Engineering and Technology, organized by Association of computer electronics and electrical engineers (ACEEE)	1	1.37
6	Conference of the International Speech Communication Association, INTERSPEECH	2	2.74
7	International Conference on Communications and Information Technology	1	1.37
8	International Conference on Advances in Computing, Communications, and Informatics (ICACCI). Institute of Electrical and Electronics Engineers	2	2.74
9	International Conference on Spoken Language and Processing	1	1.37
10	International Conference Spoken Language and Processing (ICSLP)	1	1.37
11	International Conference on Networks and Soft Computing, Institute of Electrical and Electronics Engineers	1	1.37
12	National Conference on Communication(NCC)	1	1.37
13	International Conference on Multimedia Processing Systems	1	1.37
14	Workshop on NLP for Less Privileged Languages, IJCNLP	1	1.37
15	International Conference and Development and Application Systems, Suceava, Romania	1	1.37
16	IEEE International Conference on Image and Information Processing, ICIIP, IEEE	1	1.37
17	International Conference on Language Resources and Evaluation	1	1.37
18	International Joint Conference on Neural Networks(IJCNN)	1	1.37

TABLE VI. NAME OF THE WORKSHOPS, NUMBER OF THE PAPERS, PERCENTAGE OF THE PAPERS USED IN THE SYSTEMATIC REVIEW FOR AM IN SPEECH RECOGNITION

Sl.no.	Name of the Workshops	Total papers	Workshop papers(%)
1	International Workshop on Spoken Language Technologies for Under-Resourced Languages	1	1.37
2	Workshop on Spoken Language and Technology	1	1.37
3	workshop on deep learning for speech recognition and related applications	1	1.37
4	Speech and Natural Language: Proceedings of a Workshop Held at Harriman, New York	1	1.37
5	Workshop on Speech and Natural Language. Association for Computational Linguistics	1	1.37
6	IEEE Work NNSP	1	1.37

TABLE VII. NAME OF LECTURE SERIES, NUMBER OF THE PAPERS, PERCENTAGE OF THE PAPERS USED IN THE SYSTEMATIC REVIEW FOR AM IN SPEECH RECOGNITION

Sl.no.	Name of the lecture series	Total papers	lecture series papers (%)
1	Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science). Springer Verlag	1	1.37
2	Lecture Notes Electr Eng	1	1.37

TABLE VIII. NAME OF THE UNIVERSITIES, NUMBER OF THE THESIS, PERCENTAGE OF THE THESIS USED IN THE SYSTEMATIC REVIEW FOR AM IN SPEECH RECOGNITION

Sl.no.	Name of the Universities	Number of the thesis	Percentage of the thesis(%)
1	Nanyang Technological University	1	1.37
2	Makerere University	1	1.37

B. RQ2 Aimed to Find different Types of Speech Recognition Systems used in the Study

Acoustic modeling issues were addressed for different types of SR systems in these study papers. Fig. 4 depicts the different kinds of speech recognition systems built. The speech recognition systems have been developed for isolated words, connected words, continuous speech, spontaneous speech, and multilingual speech.

The major areas of concerns were speaker-independent and dependent acoustic modeling, recognition in different noisy conditions, speech recognition for different devices, multilingual SR, recognition with weighted finite-state transducers(WFST), comparative analysis for different feature extraction techniques, recognition using subspace Gaussian mixture modeling, recognition using different subword units, and recognition for limited resource languages.

The "other" category types of the systems in Fig. 4 indicated either a combination of the methods or not explicitly mentioned. Significant research work was presented for continuous speech and connected words due to their more applications. The research findings also indicate that very little work has been reported towards spontaneous speech, conversational speech, and multilingual speech.

The reason for fewer works published for these types of systems is due to lack of resources and challenges such as context information, long conversation, and variabilities present in the environment and other conditions. It was observed that DNN based systems had been found performing better than conventional methods for these types of systems. Further multilingual SR systems are also being created by applying global phone sets.

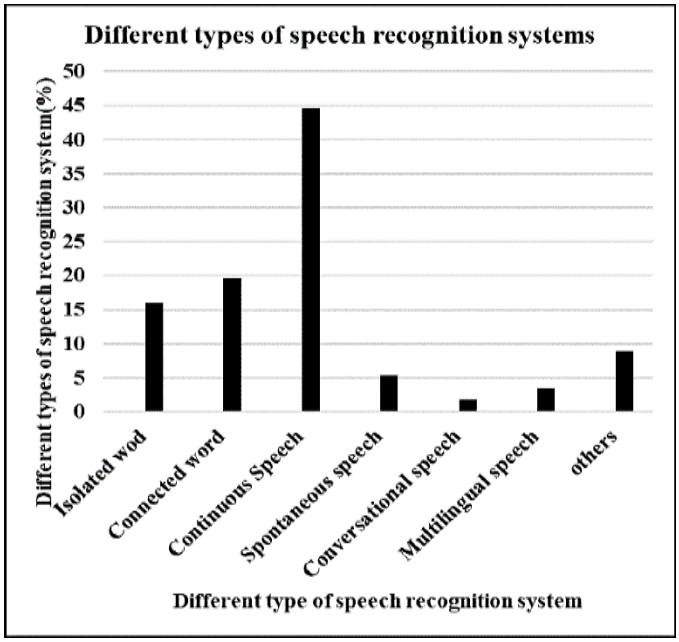

Fig. 4. Types of Speech Recognition Systems used in the Systematic Review for AM in Speech Recognition. The Category of other Types of Systems Indicated Either a Combination of the Methods or not Explicitly Mentioned.

C. RQ3 Aimed to Identify different Languages for Creating an SR Framework?

The researchers developed SR systems in different languages. Fig. 5 shows the different languages used in speech recognition. Research findings reveal that all over the world, researchers experimented for speech recognition. Different databases were developed for speech recognition. Most of the reported work belongs to the English language. It was revealed that researchers are facing problems due to a shortage of linguistic resources in SR. Multilingual speech recognition is also being experimented using a common phone set and the Global phone database.

D. RQ4 Intended to Find different Language Issues used in AM Modeling by Researchers

The studies reveal different issues about language in acoustic modeling. Language related issues are a selection of linguistic units, availability of linguistic resources, dialects, accents, contextual information, and speaker-related variabilities for acoustic modeling. It is essential to decide which language construct to use in acoustic modeling. Some languages are tonal, while others have many dialects, the acoustic models need to be generated as per the requirement. There is also a need for linguistic resources such as pronunciation dictionaries suitable for speech recognition. The researchers have used N-gram models and grammar-based rules for language modeling in speech recognition. The works also have been reported for multilingual speech recognition by developing global phone sets and speaker adaptive training.

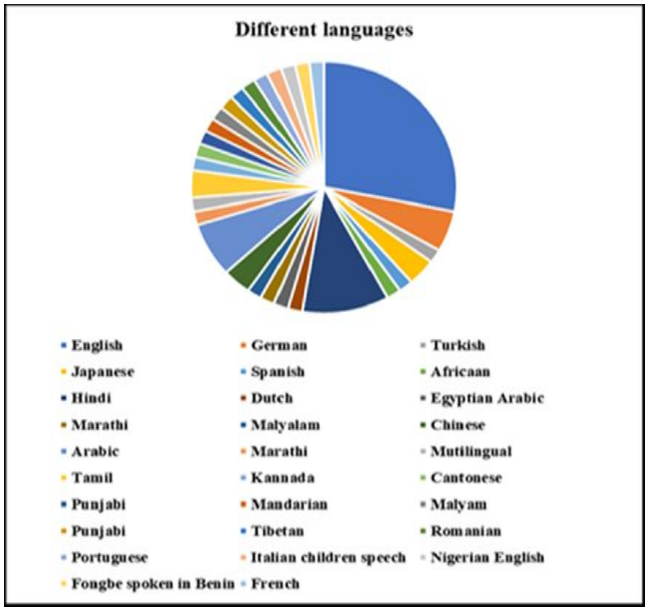

Fig. 5. Languages Applied in the Systematic Review for Acoustic Modeling in SR.

E. RQ5 Aimed to Identify different Speech Corpus used in the Study

The studies indicate that different speech corpora were used for the realization of acoustic models. Fig. 6 shows the various databases used in the systematic review study. Research outcome revels that the TIMIT speech corpus was widely used by the researchers to explore phoneme based

speech recognition as it is a well documented and phonetically balanced speech corpus with broad geographical coverage. Multilingual database GlobalPhone was used for multilingual speech recognition. It was developed with high-quality read speech. It was recorded in twenty languages with labeled data and a pronunciation dictionary. Further, the investigation also shows that mostly speech databases are available for European and American languages. Research findings also indicate that all over the world, different speech corpora in different languages were created to realize SR systems for low resource languages. Further studies also show there is a need for resources such as speech corpora and language resources for these languages. Studies also reveal that researchers developed their databases for the speech recognition systems as per their research needs.

F. RQ6 Supposed to Analyze different Software and Tools used to Experiment in the Study

The different tools used in the studied papers are Sphinx, HTK, Julius, and Kaldi for developing SR systems. Most of the research papers in the study used the HMM-based tool kit HTK. The reason for using this tool was due to well documentation and HMM-based system. HTK supports different feature extraction techniques such as MFCCs with their variants, LPCs with variants, and PLPs with variants. It also supports context-independent and context-dependent modeling. Sphinx supports MFCC and PLP speech features with delta and delta-delta features. Some expertise is needed to understand and to work on the Sphinx tool. Kaldi is being used recently in the development of speech recognition systems. It also supports DNN based methods for developing speech recognition systems. However, knowledge of shell programming and scripting in Unix/Linux based is required. Different speech processing software, such as PRATT and wave surfer, were also used. Mat Lab software was also widely used.

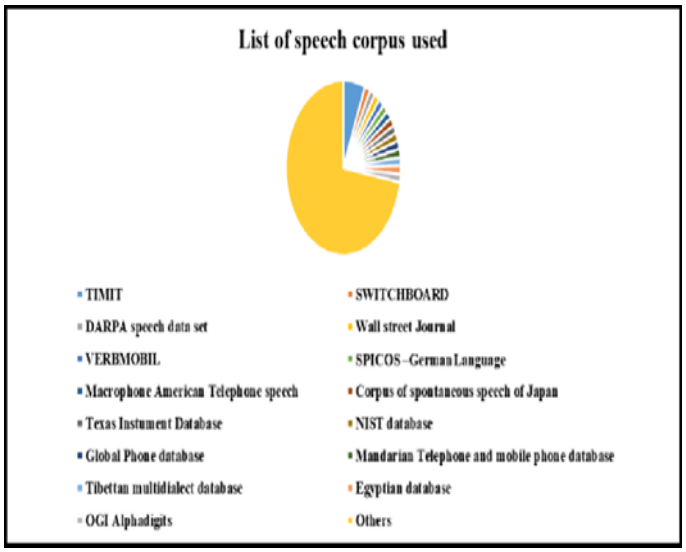

Fig. 6. Speech Corpora used in the Systematic Review for AM in Speech Recognition. The Category of other Types of Systems Indicated Self-Created Speech Corpora.

G. RQ7 Inspected different Subword Modeling Techniques are used in the Study

Research works on different subword modeling techniques were reported during the systematic review. Fig. 7 shows different subword units used in the systematic literature survey. Research findings reveal that most common sub-word acoustic models are based on the word, phonemes, syllable, and triphones. The phoneme based acoustic models have widely used in the large vocabulary continuous speech recognition system(LVCSR) system. The phonemes set are limited for any language. The phoneme based system overcome the requirement of a large number of instances. Further, phonemes are less in number; many manipulations and confusion analysis can be used. Triphone based systems were also experimented to reduce the contextual effects suffered by the phoneme based system. Context-dependent state tied triphones, crossword triphones, and word-internal triphones were used in the experiments. Syllable based system was also used instead of triphones in some studies to reduce the effect of contexts. A syllable with initial -final and onset-nucleus and coda applied for subword modeling. The category "others" in Fig. 7 shows the models used based on demisyllable, grapheme, interdigit, and character-based models.

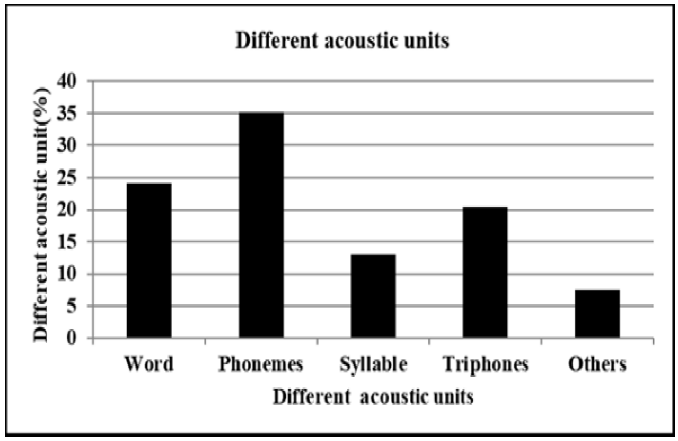

Fig. 7. Different Acoustic Modeling units used in the Systematic Review for AM in Speech Recognition. The Category of other Types of Acoustic units based on Demisyllable, Grapheme, Interdigit, and Character-based Models.

H. RQ8 Planned to Investigate the different Feature Extraction Techniques used in Acoustic Modeling

For generating acoustic models in SR, different feature extraction techniques were applied by the developers. After inspection, it was revealed that feature extraction in speech recognition was also a very much researched area in speech recognition. Researchers experimented with different feature extraction and transformation techniques to improve recognition accuracy. Fig. 8 shows different feature extraction techniques used in the systematic review. The investigations reveal that usually used feature extraction techniques are linear prediction coefficients(LPCs), Mel Frequency cepstral coefficients(MFCCs), and Perceptual linear predictive coefficients(PLPs) with their variants. Research results reveal that MFCCs are widely used coefficients. Experiments had been conducted using MFCCs with energy, first and second derivatives. Most of the research experiments were performed

with twelve MFCCs c. Some tests were also conducted using MFCCs with vocal tract area function, and power normalized cepstral coefficients(PNCC). Other feature extraction methods such as duration, intensity, mean zero-crossing, pitch, amplitude, formants, and short-time energy were also reported. Researchers also applied feature transformation techniques such as LDA and HLDA. Discriminative features were also implemented. It was observed that PLP coefficients provided better results in the case of speaker-independent speech recognition. Research works also reported vector quantization with extracted features. The advantages of vector quantization are reduced storage and reduced computation; however, the quantization error is a problem. The research findings also reveal that earlier speech recognition systems were based on time-domain processing methods, formant analysis, and linear predictive coefficients. Researchers also reported the advantages of MFCCs as good discrimination, the correlation between components, and the application of manipulation.

I. RQ9 Aimed to Find out different Classification Methods used for Systematic Reviews

Different classification methods were applied in speech recognition to develop acoustic models. Fig. 9 indicates the various classification methods utilized in this study. The commonly used classification methods are based on HMM, acoustic-phonetic approach, ANNs, dynamic time warping(DTW), Deep Neural Network(DNNs), Discriminative training, support vector machine(SVM), Fuzzy logic, CTC and Deep belief network(DBF). Research findings reveal that HMM-based systems were widely used during the past decades; however, in recent years, ANN and DNN based systems are being used. Further, research works were reported using different states, gaussian mixture models, context-independent, and context-dependent models for HMM. Discriminative training methods with objective function maximum mutual information (MMI), minimum phone error(MPE), and minimum classification error(MCE) were also applied by the researchers to improve speech recognition. Artificial neural network approaches such as Kohonen Self-organising maps, Multilayer perceptron, Time –Delay neural network, Hidden Control neural network, the combination of hidden Markov model, and connectionist probability estimators have been applied. The main strength of ANNs is their discriminative property, which is an essential property that can be used with HMMs was stated by the developers/researchers. Advantages of ANNs are the ability to learn from input data, unsupervised learning, parallel computation, system development through learning, not programming, adaptable to the environment, handling of complex interaction, and easy to use and understand. Limitations are it requires large training speech utterances and long training time.

J. RQ10 was Prepared to Find out different Performance Metrics used in the Study for SR Systems

Different quality assessment criteria used by the researchers are recognition accuracy, word correctness, word accuracy, phone error rate, frame error rate, and word error rate. Most of the searchers used word accuracy and word error rate.

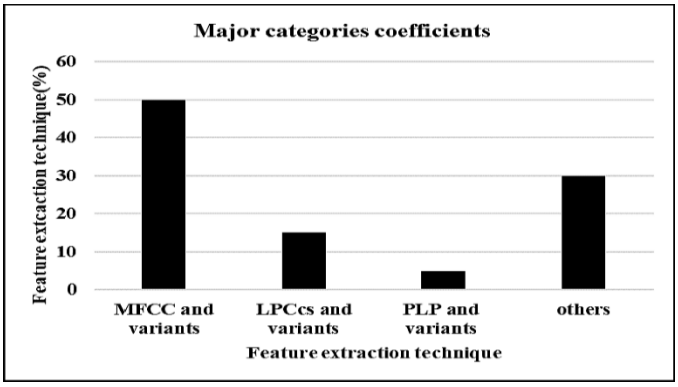

Fig. 8. Feature Extraction Methods used in the Systematic Review for Acoustic Modeling in Speech Recognition. The Category "others" Category Indicated Features other than MFCCs/LPCCs/PLPs.

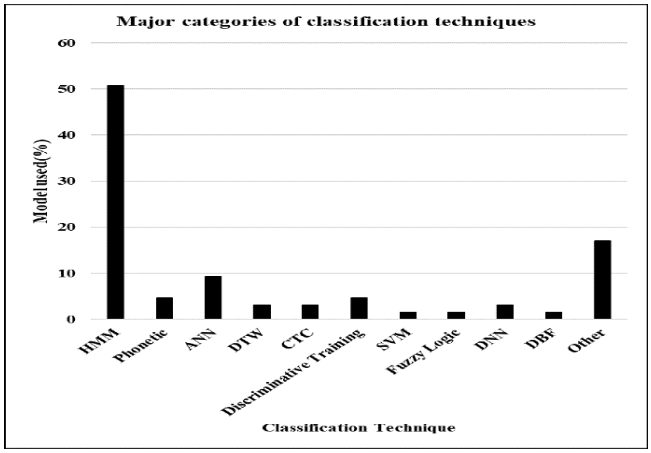

Fig. 9. The Classification Methods used in the Systematic Review of Acoustic Modeling in Speech Recognition. The Category others Include Classification Methods other than HMM/phonetic/ANN/DTW/CTC/ Discriminative Training/SVM/Fuzzy logic/DNN/DBF.

VI. DISCUSSION

Research answers to research questions were prepared after extracting the information from the finally selected papers for the systematic review process. Different research revelations have emerged from the study. It was observed that most of the research papers were provided by the IEEE library, Springer, and Science direct libraries. The conferences such as ICASSP, Eurospeech Conference on Information and Communication Technology, and INTERSPEECH are conducted explicitly for research in speech and audio processing. These conferences supplied a variety of research papers to address different problems in speech processing. It was also revealed that researchers are developing various types of speech recognition systems such as isolated words, connected words, continuous speech, spontaneous speech, and multilingual speech as per the requirement and addressed different modeling issues. Continuous speech recognition systems were widely used due to their large span of practical use.

Further, it was also observed that various acoustic modeling issues are addressed for speaker-independent, speaker-dependent acoustic modeling, different noisy conditions, speech recognition for different devices, and

multilingual speech recognition. It was also found that speech recognition is the most active research field, all over the world research community is trying to develop speech recognition systems in different languages. However, researchers are finding hardships in this field due to the unavailability of resources such as speech corpora and other linguistic resources for low resource languages. Most of the research work was reported for the English and European languages. Research outcomes also reveal that some languages such as English have systematic and well-defined speech corpuses such as TIMIT and phonetic dictionaries such as BEEP; therefore, researchers find it convenient to experiment with this standard speech corpora and dictionary. Most of the researchers are developing their resources for conducting the research work. It needs great effort in the part of these researchers to use different techniques for overcoming various constraints in this area.

Research outcomes also show that different acoustic units such as word, phoneme, syllable, character, and grapheme are being used by researchers to address issues such as related to context, data scarcity, and language modeling. Phoneme, word, triphone, and syllable based systems were generally used. The studies also reveal that phoneme based systems are widely used. Researchers are developing pronunciation dictionaries and applying language modeling techniques in speech recognition. N-gram language modeling and weighted finite-state transducers are also being used in speech recognition. Different tools and software are also being developed for acoustic modeling in speech recognition. Some of the widely used tools are HTK, Sphinx, and Kaldi. The PRATT and wave surfer were widely used for speech analysis. Matlab was also commonly used in the research.

A further area of research that was experimented extensively is feature extraction. A large number of papers have been reported by applying different feature extraction techniques to improve speech recognition. MFCCs and their variants are widely used feature extraction techniques. Further, various language issues are also being incorporated into speech recognition. Researchers also used knowledge resources in creating speech parameters.

Different classification techniques were applied to realize the different acoustic models. The commonly used classification methods are based on HMM, acoustic-phonetic approach, ANNs, dynamic time warping(DTW), Deep Neural Network(DNNs), Discriminative training, support vector machine(SVM), Fuzzy logic, CTC and Deep belief network(DBF). Research findings reveal that HMM-based systems were widely used during the past decades; however, in recent years, ANN and DNN based systems are being used. Different quality assessment criteria for measuring the performance of speech recognition are recognition accuracy, word correctness, word accuracy, phone error rate, frame error rate, and word error rate. Most of the developers used word accuracy and word error rate.

VII. Conclusion

Research questions aimed to investigate the issues regarding acoustic modeling to explore the research papers used, speech recognition system developed, languages used, language issues included, speech corpora used, acoustic modeling units applied, feature extraction techniques used, classification methods utilized, and performance metrics applied. Different quality assessment criteria were applied for the final selection of the papers. A total of seventy-three research papers were selected by applying quality assessment criteria, as mentioned in the research methodology section. The research papers have been included between 1984 to 2020 so that we attempted to include new and old researches in the field of speech recognition to understand the flow of speech recognition research in acoustic modeling. The research work started with the importance of acoustic modeling and its challenges. After that, the fundamental concept in speech recognition described understanding the acoustic modeling issues. The work presented here touched different aspects of acoustic modeling.

Research findigs show that IEEE library, Springer, and Science direct libraries provided most of the research papers. The conferences such as ICASSP, Eurospeech Conference on Information and Communication Technology, and INTERSPEECH aimed to address research papers in speech and audio processing.The investigation indicate that acoustic units such as word, phoneme, syllable, character, and grapheme were used to address context, data scarcity, and language modeling. The outcome also revealed that MFCCs, continuous speech recognition and N-gram language models were mostly used. Different classification methods have been applied. The HMM based systems were widely used for decades, but now days deep learning based systems are being experimented.Other findings also indicate that deveopers used mostly word accuracy and word error rate for the performance measurement of SR systems.

The presented research work provided deep insight into understanding different acoustic modeling issues by performing a systematic review. The outcome of the research shed light on the research flow in acoustic modeling issues and included new research areas also. The advantage of the systematic review was that research findings were revealed from the beginning, middle, and recent years of research in this field.

Research work may be extended by exploring further detailed analysis using acoustic modeling for recent techniques such as based on deep learning methods and conducting research to improve acoustic modeling in acoustic units.

Acknowledgment

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] G. I. Parisi, J. Tani, C. Weber, and S. Wermter, “Emergence of multimodal action representations from neural network self-organization,” Cogn. Syst. Res., vol. 43, pp. 208–221, Jun. 2017, doi: 10.1016/j.cogsys.2016.08.002.

[2] P. Bansal, A. Dev, and S. B. Jain, “Optimum HMM combined with vector quantization for hindi speech recognition,” IETE J. Res., vol. 54,

no. 4, pp. 239–243, 2008, doi: 10.4103/0377-2063.44216.

[3] H. Hermansky, “Perceptual linear predictive (PLP) analysis of speech,” J. Acoust. Soc. Am., vol. 87, no. 4, pp. 1738–1752, 1990, doi: 10.1121/1.399423.

[4] R. Sarikaya and J. H. L. Hansen, “ANALysis of the root-cepstrum for acoustic modeling and fast decoding in speech recognition,” EUROSPEECH 2001 - Scand. - 7th Eur. Conf. Speech Commun. Technol., pp. 687–690, 2001.

[5] S. Bhatt, A. Dev, and A. Jain, “Confusion analysis in phoneme based speech recognition in Hindi,” J. Ambient Intell. Humaniz. Comput., no. 0123456789, 2020, doi: 10.1007/s12652-020-01703-x.

[6] Y. Qi and R. A. Fox, “Analysis Of Nasal Consonants Using Perceptual Linear Prediction,” J. Acoust. Soc. Am., vol. 91, no. 3, pp. 1718–1726, 1992, doi: 10.1121/1.402451.

[7] A. Becerra, J. I. De La Rosa, and E. Gonzalez, “A case study of speech recognition in Spanish: From conventional to deep approach,” Proc. 2016 IEEE ANDESCON, ANDESCON 2016, 2017, doi: 10.1109/ANDESCON.2016.7836212.

[8] L. Bu and T. D. Chiueh, “Perceptual speech processing and phonetic feature mapping for robust vowel recognition,” IEEE Trans. Speech Audio Process., vol. 8, no. 2, pp. 105–114, 2000, doi: 10.1109/89.824695.

[9] P. Vishal B. Waghmare, “Continuous Speech Recognition System A Review,” Asian J. Comput. Sci. Inf. Technol., vol. 6, pp. 62–66, 2014, doi: 10.15520/ajcsit.v4i6.3.

[10] M. J. F. Gales, S. Watanabe, and E. Fosler-Lussier, “Structured discriminative models for speech recognition: An overview,” IEEE Signal Process. Mag., vol. 29, no. 6, pp. 70–81, Nov. 2012, doi: 10.1109/MSP.2012.2207140.

[11] R. K. Aggarwal and M. Dave, “Acoustic modeling problem for automatic speech recognition system: Advances and refinements (Part II),” Int. J. Speech Technol., vol. 14, no. 4, pp. 309–320, 2011, doi: 10.1007/s10772-011-9106-4.

[12] X. Huang and L. Deng, “An overview of modern speech recognition,” Handb. Nat. Lang. Process. Second Ed., pp. 339–366, 2010.

[13] C.-H. Lee, L. R. Rabiner, and R. Pieraccini, “Speaker Independent Continuous Speech Recognition Using Continuous Density Hidden Markov Models,” Speech Recognit. Underst., pp. 135–163, 1992, doi: 10.1007/978-3-642-76626-8_16.

[14] S. Kanthak and H. Ney, “Context-dependent acoustic modeling using graphemes for large vocabulary speech recognition,” ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc., vol. 1, no. March, 2002, doi: 10.1109/icassp.2002.5743871.

[15] C. Y. Espy-Wilson et al., “济無 No Title No Title,” Speech Commun., vol. 1, no. 1, pp. 1689–1699, 2012, doi: 10.1017/CBO9781107415324.004.D. Imseng, J. Dines, P. Motlicek, P. N. Garner, and H. Bourlard, “Comparing different acoustic modeling techniques for multilingual boosting,” 13th Annu. Conf. Int. Speech Commun. Assoc. 2012, INTERSPEECH 2012, vol. 2, pp. 1190–1193, 2012.

[16] S. Garfield, “Review of: Speech and language processing,” Cogn. Syst. Res., vol. 2, no. 2, pp. 167–172, May 2001, doi: 10.1016/s1389-0417(01)00022-5.

[17] N. Singh-miller and M. J. Collins, “Neighborhood Analysis Methods in Acoustic Modeling for Automatic Speech Recognition by X7 by,” Electr. Eng., 2010.

[18] Y. Gong and J. Haton, “Issues in Acoustic Modeling of Speech for Automatic Speech Recognition Issues in acoustic modeling of speech for automatic speech recognition apport de recherche,” no. January, 1994.

[19] S. Tasnim Swarna, S. Ehsan, S. Islam, and M. E. Jannat, “A Comprehensive Survey on Bengali Phoneme Recognition,” in Proceedings of the International Conference on Engineering Research, Innovation and Education 2017 ICERIE 2017, 2017, vol. 13, no. 15, pp. 1–7.

[20] H. Yao, M. An, J. Xu, and J. Liu, “Efficient Acoustic Modeling Method for Unsupervised Speech Recognition using Multi-Task Deep Neural Network,” in 4th National Conference on Electrical, Electronics and Computer Engineering (NCEECE 2015) Efficient, 2016, pp. 365–370, doi: 10.2991/nceece-15.2016.72.

[21] H. Nanjo, K. Kato, and T. Kawahara, “Speaking rate dependent acoustic modeling for spontaneous lecture speech recognition,” EUROSPEECH 2001 - Scand. - 7th Eur. Conf. Speech Commun. Technol., no. May, pp. 2531–2534, 2001.

[22] Z. Wu and Z. Cao, “Improved MFCC-based feature for robust speaker identification,” Tsinghua Sci. Technol., vol. 10, no. 2, pp. 158–161, 2005, doi: 10.1016/S1007-0214(05)70048-1.

[23] G. Huang and M. J. Er, “Model-based articulatory phonetic features for improved speech recognition,” Proc. Int. Jt. Conf. Neural Networks, 2012, doi: 10.1109/IJCNN.2012.6252748.

[24] W. Zou, D. Jiang, S. Zhao, G. Yang, and X. Li, “A comparable study of modeling units for end-to-end Mandarin speech recognition,” 2018 11th Int. Symp. Chinese Spok. Lang. Process. ISCSLP 2018 - Proc., pp. 369–373, 2018, doi: 10.1109/ISCSLP.2018.8706661.

[25] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, “Systematic literature reviews in software engineering - A systematic literature review,” Inf. Softw. Technol., vol. 51, no. 1, pp. 7–15, 2009, doi: 10.1016/j.infsof.2008.09.009.

[26] K. Sharma and S. Bhatt, “SQL injection attacks - a systematic review,” Int. J. Inf. Comput. Secur., vol. 11, no. 4/5, p. 493, 2019, doi: 10.1504/ijics.2019.101937.

[27] A. Kumar and G. Garg, “Systematic literature review on context-based sentiment analysis in social multimedia,” Multimed. Tools Appl., 2019, doi: 10.1007/s11042-019-7346-5.

[28] R. Malhotra, “A systematic review of machine learning techniques for software fault prediction,” Appl. Soft Comput. J., vol. 27, pp. 504–518, 2015, doi: 10.1016/j.asoc.2014.11.023.

[29] A. B. Nassif, I. Shahin, I. Attili, M. Azzeh, and K. Shaalan, “Speech Recognition Using Deep Neural Networks: A Systematic Review,” IEEE Access, vol. 7, pp. 19143–19165, 2019, doi: 10.1109/ACCESS.2019.2896880.

[30] C. H. Lee, L. R. Rabiner, R. Pieraccini, and J. G. Wilpon, “Acoustic modeling of subword units for speech recognition,” ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc., vol. 2, no. September 2014, pp. 721–724, 1990, doi: 10.1109/icassp.1990.115885.

[31] S. Bhatt, A. Dev, and A. Jain, “Hidden Markov Model Based Speech Recognition-A Review,” in International Conference on "Computing for Sustainable Global Development India Com IEEE Conference, 2018, pp. 3367–3372.

[32] J. Li, L. Deng, Y. Gong, and R. Haeb-Umbach, “An overview of noise-robust automatic speech recognition,” IEEE Transactions on Audio, Speech and Language Processing, vol. 22, no. 4. Institute of Electrical and Electronics Engineers Inc., pp. 745–777, 2014, doi: 10.1109/TASLP.2014.2304637.

[33] A. Waris and R. K. Aggarwal, “Acoustic modeling in Automatic Speech Recognition - A Survey,” Proc. 2nd Int. Conf. Electron. Commun. Aerosp. Technol. ICECA 2018, no. Iceca, pp. 1408–1412, 2018, doi: 10.1109/ICECA.2018.8474889.

[34] M. Cutajar, E. Gatt, I. Grech, O. Casha, and J. Micallef, “Comparative study of automatic speech recognition techniques,” IET Signal Process., vol. 7, no. 1, pp. 25–46, 2013, doi: 10.1049/iet-spr.2012.0151.

[35] J. Padmanabhan and M. J. J. Premkumar, “Machine learning in automatic speech recognition: A survey,” IETE Tech. Rev. (Institution Electron. Telecommun. Eng. India), vol. 32, no. 4, pp. 240–251, 2015, doi: 10.1080/02564602.2015.1010611.

[36] R. K. Aggarwal and M. Dave, “Acoustic modeling problem for automatic speech recognition system: Conventional methods (Part I),” Int. J. Speech Technol., vol. 14, no. 4, pp. 297–308, 2011, doi: 10.1007/s10772-011-9108-2.

[37] J. Wu, E. Yılmaz, M. Zhang, H. Li, and K. C. Tan, “Deep Spiking Neural Networks for Large Vocabulary Automatic Speech Recognition,” Front. Neurosci., vol. 14, p. 199, Mar. 2020, doi: 10.3389/fnins.2020.00199.

[38] A. Kumar and R. K. Aggarwal, “A time delay neural network acoustic modeling for hindi speech recognition,” in Lecture Notes in Networks and Systems, vol. 94, Springer, 2020, pp. 425–432.

[39] J. Ludeña-Choez and A. Gallardo-Antolín, “NMF-based spectral

analysis for acoustic event classification tasks," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 7911 LNAI, no. June, pp. 9–16, 2013, doi: 10.1007/978-3-642-38847-7.

[40] D. Jurafsky and J. H. Martin, "Speech recognition: advanced topics," Speech Lang. Process. An Introd. to Nat. Lang. Process. Comput. Linguist. Speech Recognit., pp. 1–34, 2007.

[41] S. F. Chen and J. Goodman, "An Empirical Study of Smoothing Techniques for Language Modeling," 1996.

[42] M. Sundermeyer, R. Schlüter, and H. Ney, "LSTM Neural Networks for Language Modeling."

[43] G. Salvi, "Developing Acoustic Models for Automatic Speech Recognition in Swedish," Eur. Student J. Lang. Speech, no. June 1999, pp. 1–16, 1999.

[44] S. Bhatt, A. Jain, and A. Dev, "CICD acoustic modeling based on Monophone and Triphone for HINDI Speech Recognition," in International Conference on Artificial Intelligence and Speech Technology (AIST2019) 14-15th November, 2019.

[45] I. Szöke, L. Burget, J. Černocký, and M. Fapšo, "Sub-word modeling of out of vocabularywords in spoken term detection," 2008 IEEE Work. Spok. Lang. Technol. SLT 2008 - Proc., no. 4, pp. 273–276, 2008, doi: 10.1109/SLT.2008.4777893.

[46] S. Karpagavalli and E. Chandra, "A Review on Sub-word unit Modeling in Automatic Speech Recognition," IOSR J. VLSI Signal Process., vol. 6, no. 6, pp. 77–84, 2016, doi: 10.9790/4200-0606017784.

[47] J. Mehler, J. Y. Dommergues, U. Frauenfelder, and J. Segui, "The syllable's role in speech segmentation," J. Verbal Learning Verbal Behav., vol. 20, no. 3, pp. 298–305, 1981, doi: 10.1016/S0022-5371(81)90450-3.

[48] C. M. White, S. Khudanpur, and J. K. Baker, "An investigation of acoustic models for multilingual code-switching," Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH, pp. 2691–2694, 2008.

[49] M. Agarwal et al., "MULTILINGUAL ACOUSTIC MODELING FOR SPEECH RECOGNITION BASED ON SUBSPACE GAUSSIAN MIXTURE MODELS Luk ´ a s Brno University of Technology , Czech Republic , { burget , schwarzp } @ fit . vutbr . cz ; IIIT Allahabad , India ; 3 Bo˘ gazic," pp. 4334–4337, 2010.

[50] M. A. Anusuya and S. K. Katti, "1001.2267," Int. J. Comput. Sci. Inf. Secur., vol. 6, no. 3, pp. 181–205, 2009.

[51] C. Y. Fook, H. Muthusamy, L. S. Chee, S. Bin Yaacob, and A. H. B. Adom, "Comparison of speech parameterization techniques for the classification of speech disfluencies," Turkish J. Electr. Eng. Comput. Sci., vol. 21, no. SUPPL. 1, pp. 1983–1994, 2013, doi: 10.3906/elk-1112-84.

[52] C. D. Shulby, M. D. Ferreira, R. F. de Mello, and S. M. Aluisio, "Theoretical learning guarantees applied to acoustic modeling," J. Brazilian Comput. Soc., vol. 25, no. 1, p. 1, Dec. 2019, doi: 10.1186/s13173-018-0081-3.

[53] C. Y. Espy-Wilson, "An Acoustic-Phonetic Approach to Speech Recognition : Application to the Semivowels," no. 531, 1987.

[54] S. Chang, L. Shastri, and S. Greenberg, "Automatic phonetic transcription of spontaneous speech (American English)," 6th Int. Conf. Spok. Lang. Process. ICSLP 2000, 2000.

[55] G. Saon and J. T. Chien, "Large-vocabulary continuous speech recognition systems: A look at some recent advances," IEEE Signal Process. Mag., vol. 29, no. 6, pp. 18–33, 2012, doi: 10.1109/MSP.2012.2197156.

[56] CMUSphinx, "CMUSphinx Open Source Speech Recognition." 2019.

[57] "(No Title)." [Online]. Available: https://www.imbs.uci.edu/files/docs/2009/Deng.pdf. [Accessed: 03-Mar-2020].

[58] S. Kim and F. Metze, "Acoustic-to-Word Models with Conversational Context Information."

[59] J. Pan, C. Liu, Z. Wang, Y. Hu, and H. Jiang, "Investigation of Deep Neural Networks (DNN) for large vocabulary continuous speech recognition: Why DNN surpasses GMMS in acoustic modeling," 2012 8th Int. Symp. Chinese Spok. Lang. Process. ISCSLP 2012, pp. 301–305, 2012, doi: 10.1109/ISCSLP.2012.6423452.

[60] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering – A systematic literature review," 2008, doi: 10.1016/j.infsof.2008.09.009.

[61] L. R. Rabiner, "On the Application of Energy Contours to the Recognition of Connected Word Sequences," AT&T Bell Lab. Tech. J., vol. 63, no. 9, pp. 1981–1995, Nov. 1984, doi: 10.1002/j.1538-7305.1984.tb00085.x.

[62] M. A. Bush and G. E. Kopec, "Network-based connected digit recognition," IEEE Trans. Acoust., vol. 35, no. 10, pp. 1401–1413, 1987, doi: 10.1109/TASSP.1987.1165057.

[63] L. Kai-fu and H. Hsiao-Wuen, "Speaker-Independent Phone Recognition Using Hidden Markov Models," IEEE Trans. Acoust. Speech Signal Process., vol. 31, no. 11, pp. 1641–1648, 1989.

[64] J. Picone, "Continuous Speech Recognition Using Hidden Markov Models," IEEE ASSP Mag., vol. 7, no. 3, pp. 26–41, 1990, doi: 10.1109/53.54527.

[65] K. F. Lee, "Context-Dependent Phonetic Hidden Markov Models for Speaker-Independent Continuous Speech Recognition," IEEE Trans. Acoust., vol. 38, no. 4, pp. 599–609, 1990, doi: 10.1109/29.52701.

[66] U. Dağitan and N. Yalabik, "Connected Word Recognition Using Neural Networks," in Neurocomputing, Springer Berlin Heidelberg, 1990, pp. 297–300.

[67] Y. Normandin, R. Cardin, and R. De Mori, "High-Performance Connected Digit Recognition Using Maximum Mutual Information Estimation," IEEE Trans. Speech Audio Process., vol. 2, no. 2, pp. 299–311, 1994, doi: 10.1109/89.279279.

[68] C. C. Sekhar and B. Yegnanarayana, "Recognition of Stop-Consonant-Vowel (SCV) Segments in Continuous Speech using Neural Network Models," IETE J. Res., vol. 42, no. 4–5, pp. 269–280, 1996, doi: 10.1080/03772063.1996.11415933.

[69] C. Nieuwoudta and E.C. Bothab, "Connected digit recognition in Afrikaans using hidden Markov models," 1999.

[70] T. Pruthi, S. Saksena, and P. K. Das, "Swaranjali: Isolated word recognition for Hindi language using VQ and HMM," Int. Conf. Multimed. Process. Syst., pp. 13–15, 2000.

[71] J. Ben, W. G. Wan, and X. Q. Yu, "Phoneme based speaker-independent english command recognition," J. Shanghai Univ., vol. 7, no. 2, pp. 163–167, 2003, doi: 10.1007/s11741-003-0085-9.

[72] A. Dev, S. S. Agrawal, and D. R. Choudhury, "Categorization of Hindi phonemes by neural networks," AI Soc., vol. 17, no. 3–4, pp. 375–382, 2003, doi: 10.1007/s00146-003-0263-0.

[73] M. Gerosa, D. Giuliani, and F. Brugnara, "Acoustic variability and automatic recognition of children's speech," Speech Commun., vol. 49, no. 10–11, pp. 847–860, 2007, doi: 10.1016/j.specom.2007.01.002.

[74] M. M. Azmi, H. Tolba, S. Mahdy, and M. Fashal, "Syllable-based automatic Arabic speech recognition in noisy-telephone channel," WSEAS Trans. Signal Process., vol. 4, no. 4, pp. 211–220, 2008.

[75] C. Kurian and K. Balakrishnan, "Development & evaluation of different acoustic models for Malayalam continuous speech recognition," Procedia Eng., vol. 30, no. 2011, pp. 1081–1088, 2012, doi: 10.1016/j.proeng.2012.01.966.

[76] Z. He and Z. Liu, "Chinese connected word speech recognition based on derivative dynamic time warping," in Advanced Materials Research, 2012, vol. 542–543, pp. 1324–1329, doi: 10.4028/www.scientific.net/AMR.542-543.1324.

[77] K. Kumar, R. K. Aggarwal, and A. Jain, "A Hindi speech recognition system for connected words using HTK," Int. J. Comput. Syst. Eng., vol. 1, no. 1, p. 25, 2012, doi: 10.1504/ijcsyse.2012.044740.

[78] R. K. Aggarwal and M. Dave, "Integration of multiple acoustic and language models for improved Hindi speech recognition system," Int. J. Speech Technol., vol. 15, no. 2, pp. 165–180, 2012, doi: 10.1007/s10772-012-9131-y.

[79] M. Borsky, P. Pollak, and P. Mizera, "Advanced acoustic modelling techniques in MP3 speech recognition," Eurasip J. Audio, Speech, Music Process., vol. 2015, no. 1, pp. 2–7, 2015, doi: 10.1186/s13636-015-0064-7.

[80] E. Zarrouk and Y. Benayed, "Hybrid SVM/HMM Model for the Arab Phonemes Recognition," Int. Arab J. Inf. Technol., vol. 13, no. 5, pp.

574–582, 2016.

[81] K. M. O. Nahar, M. Abu Shquier, W. G. Al-Khatib, H. Al-Muhtaseb, and M. Elshafei, “Arabic phonemes recognition using hybrid LVQ/HMM model for continuous speech recognition,” Int. J. Speech Technol., vol. 19, no. 3, pp. 495–508, 2016, doi: 10.1007/s10772-016-9337-5.

[82] M. K. Khwaja, P. Vikash, P. Arulmozhivarman, and S. Lui, “Robust phoneme classification for automatic speech recognition using hybrid features and an amalgamated learning model,” Int. J. Speech Technol., vol. 19, no. 4, pp. 895–905, 2016, doi: 10.1007/s10772-016-9377-x.

[83] P. Mittal and N. Singh, “Development and analysis of Punjabi ASR system for mobile phones under different acoustic models,” Int. J. Speech Technol., vol. 22, no. 1, pp. 219–230, 2019, doi: 10.1007/s10772-019-09593-x.

[84] Y. Zhao, J. Yue, X. Xu, L. Wu, and X. Li, “End-to-End-Based Tibetan Multitask Speech Recognition,” IEEE Access, vol. 7, pp. 162519–162529, 2019, doi: 10.1109/ACCESS.2019.2952406.

[85] S. Bhatt, A. Dev, and A. Jain, “Effects of the Dynamic and Energy based Feature Extraction on Hindi Speech Recognition,” Recent Adv. Comput. Sci. Commun., vol. 13, 2020, doi: 10.2174/2213275912666191001215916.

[86] V. Kadyan and M. Kaur, “SGMM-Based Modeling Classifier for Punjabi Automatic Speech Recognition System,” in Advances in Intelligent Systems and Computing, 2020, vol. 767, pp. 149–155, doi: 10.1007/978-981-13-9680-9_12.

[87] H. Bourlard, Y. Kamp, and C. J. Wellekens, “SPEAKER DEPENDENT CONNECTED SPEECH RECOGNITION VIA PHONEMIC MARKOV MODELS.,” in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 1985, pp. 1213–1216, doi: 10.1109/icassp.1985.1168285.

[88] D. Kenny, P., Parthasarathy, S., Gupta, V. N., Lennig, M., Mermelstein, P., & O’Shaughnessy, “Energy, Duration and Markov Models,” in In Second European Conference on Speech Communication and Technology., 1991, no. September, pp. 927–930.

[89] F. J. Caminero-Gil and C. M. D. Torre-Munilla, C. D. L., Hernandez-Gomez, L., & Álamo, “New N-best based rejection techniques for improving a real-time telephonic connected word recognition system.,” in 4th European Conference on Speech Communication and Technology EUROSPEECH ’95, 1995, no. September, pp. 2099–2102.

[90] T. Schultz and I. Rogina, “Acoustic and language modeling of human and nonhuman noises for human-to-human spontaneous speech recognition,” ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc., vol. 1, no. June 1995, pp. 293–296, 1995, doi: 10.1109/icassp.1995.479531.

[91] L. Fissore and F.Ravera and P. Laface, “Acoustic-phonetic modeling for Flexible vocabulary speech recognition,” 1995.

[92] A. Anastasakos, R. Schwartz, and H. Shu, “Duration modeling in large vocabulary speech recognition,” in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 1995, vol. 1, pp. 628–631, doi: 10.1109/icassp.1995.479676.

[93] E. Bonafonte Cávez, A., Estany, R., & Vives, “Study of subword units for Spanish speech recognition.,” in In Proceedings of the 4th EUROPEAN CONFERENCE ON SPEECH COMMUNICATION AND TECHNOLOGY ESCA-JM PARDO, E. ENRIQUEZ, J. ORTEGA, J. FERREIROS GTM-UPM., 1995, pp. 1607–1610.

[94] K. Beulen, E. Bransch, and H. Ney, “State tying for context dependent phoneme models,” Fifth Eur. Conf., no. August 2002, pp. 3–6, 1997.

[95] Y. H. Kao and L. Netsch, “Inter-digit HMM connected digit recognition using the macrophone corpus,” in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 1997, vol. 3, pp. 1739–1742, doi: 10.1109/icassp.1997.598860.

[96] D. Sima, M., Croitoru, V., & Burileanu, “Performance analysis on speech recognition using neural networks.,” in In Proceedings of the International Conference and Development and Application Systems, Suceava, Romania, 1998, pp. 259–266.

[97] W. Byrne et al., “Towards language independent acoustic modeling,” ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc., vol. 2, no. 1, pp. 1029–1032, 2000, doi: 10.1109/ICASSP.2000.859138.

[98] N. Mukherjee, N. Rajput, L. V. Subramaniam, and A. Verma, “On deriving a phoneme model for a new language,” 6th Int. Conf. Spok. Lang. Process. ICSLP 2000, pp. 8–11, 2000.

[99] G. Stemmer and N. Elmar, “Acoustic Modeling of Foreign Words in a German Speech Recognition System,” in Acoustic modeling of foreign words in a German speech recognition system. In Seventh European Conference on Speech Communication and Technology., 2001, pp. 1–4.

[100] M. Magimai-Doss, S. Bengio, and H. Bourlard, “Joint decoding for phoneme-grapheme continuous speech recognition,” in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2004, vol. 1, doi: 10.1109/icassp.2004.1325951.

[101] W. Macherey, L. Haferkamp, R. Schl, and H. Ney, “Investigations on Error Minimizing Training Criteria for Discriminative Training in Automatic Speech Recognition,” in In Ninth Euro-pean Confer-ence on Speech Communica-tion and Tech-nology, 2005, pp. 2133–2136.

[102] A. Lakshmi and H. A. Murthy, “A syllable based continuous speech recognizer for Tamil,” INTERSPEECH 2006 9th Int. Conf. Spok. Lang. Process. INTERSPEECH 2006 - ICSLP, vol. 4, pp. 1878–1881, 2006.

[103] M. D. R.K. Aggarwal, “Implementing a Speech Recognition System Interface for Indian Languages - ACL Anthology,” in Proceedings of the IJCNLP-08 Workshop on NLP for Less Privileged Languages, 2008.

[104] S. Amuda, H. Bořil, A. Sangwan, and J. H. L. Hansen, “Limited resource speech recognition for Nigerian English,” in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2010, no. January, pp. 5090–5093, doi: 10.1109/ICASSP.2010.5495036.

[105] V. Patil and P. Rao, “Acoustic features for detection of aspirated stops,” 2011 Natl. Conf. Commun. NCC 2011, 2011, doi: 10.1109/NCC.2011.5734735.

[106] N. Hammami, M. Bedda, and F. Nadir, “The second-order derivatives of MFCC for improving spoken Arabic digits recognition using tree distributions approximation model and HMMs,” in International Conference on Communications and Information Technology - Proceedings, 2012, pp. 1–5, doi: 10.1109/ICCITechnol.2012.6285769.

[107] S. Sinha, S. S. Agrawal, and A. Jain, “Continuous density Hidden Markov Model for context dependent Hindi speech recognition,” Proc. 2013 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2013, pp. 1953–1958, 2013, doi: 10.1109/ICACCI.2013.6637481.

[108] S. Tripathy, N. Baranwal, and G. C. Nandi, “A MFCC based Hindi speech recognition technique using HTK Toolkit,” 2013 IEEE 2nd Int. Conf. Image Inf. Process. IEEE ICIIP 2013, no. December, pp. 539–544, 2013, doi: 10.1109/ICIIP.2013.6707650.

[109] A. Chaudhary, M. R. Chauhan, and M. G. Gupta, “Automatic speech recognition system for isolated and connected words of Hindi language by using hidden markov model toolkit (HTK),” in Proc. of Int. Conf. on Emerging Trends in Engineering and Technology, organized by Association of computer electronics and electrical engineers (ACEEE), 2013, pp. 847–853, doi: DOI: 03.AETS.2013.3.234.

[110] P. P. Patil and S. A. Pardeshi, “Marathi connected word speech recognition system,” in 1st International Conference on Networks and Soft Computing, ICNSC 2014 - Proceedings, 2014, pp. 314–318, doi: 10.1109/CNSC.2014.6906687.

[111] N. T. Vu, D. Imseng, D. Povey, P. Motlicek, T. Schultz, and H. Bourlard, “Multilingual deep neural network based acoustic modeling for rapid language adaptation,” ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc., pp. 7639–7643, 2014, doi: 10.1109/ICASSP.2014.6855086.K. S. Akhila and R. Kumaraswamy, “Comparative analysis of Kannada phoneme recognition using different classifiers,” in International Conference on Trends in Automation, Communication and Computing Technologies, I-TACT 2015, 2016, doi: 10.1109/ITACT.2015.7492683.

[112] T. Wong et al., “Syllable based DNN-HMM Cantonese Speech-to-Text System,” in the Tenth International Conference on Language Resources and Evaluation, 2016, pp. 3856–3862.

[113] A. Kaur and A. Singh, “Optimizing feature extraction techniques constituting phone based modelling on connected words for Punjabi automatic speech recognition,” in 2016 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2016, 2016, pp. 2104–2108, doi: 10.1109/ICACCI.2016.7732362.

[114] P. Ghahremani, H. Hadian, H. Lv, D. Povey, and S. Khudanpur, "Acoustic modeling from frequency-domain representations of speech," Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH, vol. 2018-Septe, pp. 1596–1600, 2018, doi: 10.21437/Interspeech.2018-1453.

[115] Y. Zhao, L. Dong, S. Xu, and B. Xu, "Syllable-Based Acoustic Modeling with CTC for Multi-Scenarios Mandarin speech recognition," Proc. Int. Jt. Conf. Neural Networks, vol. 2018-July, no. 2016, 2018, doi: 10.1109/IJCNN.2018.8489589.

[116] D. R. Liu, K. Y. Chen, H. Y. Lee, and L. S. Lee, "Completely unsupervised phoneme recognition by adversarially learning mapping relationships from audio embeddings," in Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2018, vol. 2018-Septe, no. September, pp. 3748–3752, doi: 10.21437/Interspeech.2018-1800.

[117] V. Digalakis, M. Ostendorf, and J. R. Rohlicek, "Improvements in the stochastic segment model for Phoneme recognition," in Proceedings of the workshop on Speech and Natural Language. Association for Computational Linguistics, 1989., 1989, pp. 332–338, doi: 10.3115/1075434.1075491.

[118] S. Austin et al., "BBN real-time speech recognition demonstrations," in In Speech and Natural Language: Proceedings of a Workshop Held at Harriman, New York, 1992, pp. 250–251, doi: 10.3115/1075527.1075584.

[119] A. Biem, S. Katagiri, and B. H. Juang, "Discriminative feature extraction for speech recognition," Neural Networks Signal Process. III - Proc. 1993 IEEE Work. NNSP 1993, pp. 392–401, 1993, doi: 10.1109/NNSP.1993.471849.

[120] A.-R. Mohamed, G. Dahl, and G. Hinton, "Deep Belief Networks for Phone Recognition," in In Nips workshop on deep learning for speech recognition and related applica-tions, 2009, vol. 4, no. 5, pp. 1–9, doi: 10.4249/scholarpedia.5947.

[121] S. Bhatt, A. Dev, and A. Jain, "Hindi Speech Vowel Recognition Using Hidden Markov Model," in The 6th Intl. Workshop on Spoken Language Technologies for Under-Resourced Languages , 2018, vol. 1, pp. 196–199.

[122] D. Caseiro and I. Trancoso, "Large vocabulary continuous speech recognition using weighted finite-state transducers," in Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science), 2002, vol. 2389, pp. 91–99, doi: 10.1007/3-540-45433-0_15.

[123] M. Maseri and M. Mamat, "Malay language speech recognition for preschool children using hidden markov model (HMM) system Training," Lect. Notes Electr. Eng., vol. 481, no. ii, pp. 205–214, 2019, doi: 10.1007/978-981-13-2622-6_21.

[124] M. Jackson, "Automatic Speech Recognition : Human Computer Interface for Kinyarwanda Language," 2005.

[125] DO VAN HAI, "Acoustic Modeling for Speech Recognition under Limited Training Data Conditions," Nanyang Technological University, 2015.