CHAPTER 2

REVIEW OF LITERATURE

During recent years, ASR has started to emerge as a pragmatic and useful technology for various applications. During the last two decades, a remarkable advancement has been achieved in the field of ASR Technology. In this chapter, the history of an ASR system and the related works carried out in ASR and significant findings from various authors are briefly discussed. For a comprehensive understanding, the review of literature is presented under six categories, which are as follows:

- Speech Recognition for English and other Similar Languages,
- Multilingual Speech Recognition,
- Speech Recognition using Hybrid Techniques,
- Speech Recognition for Tamil Language,
- Noisy Speech Recognition, and
- Speech Signal Enhancement Techniques for Noisy Speech Recognition.

2.1 Historical Background of an ASR System

The general speech recognition research was initiated in the early 1870s by Alexander Graham Bell, who discovered the solution for converting sound waves into electrical impulses through telephone. Later, more mathematical and scientific techniques have been developed for interpreting human language. Davis, K. H., et al. (1952) have demonstrated an isolated digit recognition system for a single speaker. Further attempt was carried out by IBM Corporation and Advanced Research Projects Agency (ARPA) speech understanding research in 1960s and 1970s. During 1980s, IBM’s Dragon naturally speaking systems were developed for creating text documents through voice dictation.

In the year 1990, great progress was achieved in the development of software tools that enabled many individual research programmes all over the world. The Hidden Markov Model Tool Kit (HTK) developed by the Cambridge
University team is one of the most widely adopted software tools by all the leading speech research companies (Young, S.J, 1994). HTK toolkit has been the state-of-the-art ever since in the ASR research.

During recent years, numerous benefits have been attained by the development of ASR application. Today, Large Vocabulary Continuous Speech Recognition (LVCSR) system is available for creating text documents by using voice input. Speech recognition is also being used with robots to perform simple operations based on the voice commands. The most recent and popular use of ASR today is the iPhone with Siri programme. Siri is an intelligent, voice-activated, personal assistant and knowledge navigator, which uses a natural language user interface to support general query system with web services (Tyler Fulcher, 2012). Table 2.1 shows some of the major significant progresses of an ASR research.

**TABLE 2.1**

**Significant Progress of an ASR Research**

<table>
<thead>
<tr>
<th>Year</th>
<th>Details of Research Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>1952</td>
<td>First speaker dependent, isolated digit recognizer</td>
</tr>
<tr>
<td>1968</td>
<td>Speech recognition using DTW</td>
</tr>
<tr>
<td>1970</td>
<td>HMM for speech recognition</td>
</tr>
<tr>
<td>1971</td>
<td>Real time word recognition system using HMM</td>
</tr>
<tr>
<td>1972</td>
<td>Isolated word recognition</td>
</tr>
<tr>
<td>1973</td>
<td>Connected word recognition</td>
</tr>
<tr>
<td>1975</td>
<td>Speech and character recognition application</td>
</tr>
</tbody>
</table>
| 1976 | HARPY-Connected Speech Recognition System  
Utterance classification  
Continuous speech recognition with HMM  
Speech recognition for Artificial Intelligence (AI) |
<p>| 1978 | Speaker independent Isolated Word Recognition |
| 1982 | Low cost speaker dependent speech recognition with Walsh Hadamard Transform |
| 1983 | Optimization with empirical bayes approach for speech recognition |</p>
<table>
<thead>
<tr>
<th>Year</th>
<th>Details of Research Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>1984</td>
<td>Turing good formulas for probabilities of unseen events</td>
</tr>
<tr>
<td>1986</td>
<td>Digit recognition with energy and filterbanks features Rabiner Markov models to Speech Recognition</td>
</tr>
<tr>
<td>1989</td>
<td>Probabilistic mixer model for noisy speech recognition</td>
</tr>
<tr>
<td>1991</td>
<td>Speech recognition using Context Free Grammars</td>
</tr>
<tr>
<td>1992</td>
<td>Spoken dialogue system for German sentences using Bayesian rule</td>
</tr>
<tr>
<td>1994</td>
<td>Connectionist network integrated with HMM</td>
</tr>
<tr>
<td>1998</td>
<td>TI digits recognition with formant frequencies</td>
</tr>
<tr>
<td>2000</td>
<td>Emotion speech recognition</td>
</tr>
</tbody>
</table>
| 2001 | - Julius—An Open Source Real-Time Large Vocabulary Recognition Engine  
- Generalized confidence score for speech recognition and utterance verification  
- GMM for accent identification for mandarin speech  
- HMM with formant features in spectral domain |
| 2002 | Stanford Research Institute Language Modeling (SRILM) Toolkit |
| 2003 | Speaker identification using spectral information and stylistic features |
| 2004 | ASR using Dynamic Basiyen Network |
| 2006 | Robot audition system with missing features and Voice Activity Detection (VAD) |
| 2008 | Context dependent quantization |
| 2009 | Adaptive speech enhancement and speech recognition with normalization techniques |
| 2010 | - Effective frame selection approach based on SNR and energy distance  
- Browsing meeting recordings by speaker and keyword with Graphical User Interface (GUI)  
- Non-native mandarin speech recognition using Maximum Likelihood Linear Regression (MLLR), Maximum a Posteriori (MAP)  
- Visual speech recognition for Chinese words  
- Multi-stream auditory features for Audio-Visual Speech Recognition (AVSR) |
<p>| 2011 | Distributed Turkish CSR using packet loss concealment |
| 2012 | Deep Neural Networks for Acoustic Modeling in Speech Recognition |
| 2013 | Deep Learning approaches |</p>
<table>
<thead>
<tr>
<th>Year</th>
<th>Details of Research Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014</td>
<td>Siri for Android</td>
</tr>
</tbody>
</table>

To date, the need for an ASR system in real world speech oriented applications, that are accurate and robust to environment and speaker constraints, are not fulfilled. Substantial efforts have been made by many researchers to develop an ASR system that can mimic the human performance. Nevertheless, computer based systems do not yet possess the capability and flexibility of understanding all speech in any acoustic environment by any individual. Most commercial companies claim that recognition software can achieve 98 to 99% accuracy, if the system is operated under optimal conditions.

An ultimate goal of ASR is to make a real time system to achieve 100% accuracy in recognizing all words which are independent of vocabulary, speaker and environment characteristics. But the research is not close enough to meet the human ability ([http://www.docsoft.com/resources/Studies/Whitepapers/whitepaper-ASR.pdf](http://www.docsoft.com/resources/Studies/Whitepapers/whitepaper-ASR.pdf)). Various techniques have been suggested by many researchers for developing different applications. They are presented below:

### 2.2 Related Works on ASR for English and other Similar Languages

Reichert, J. et al. (1999) have presented the efforts in developing a speaker independent LVCSR engine for Mandarin Chinese using GlobalPhone multilingual database. Two pass approaches have originated in which Pinyin hypotheses are generated first and then they are transformed into Chinese character hypotheses. Phoneme Vs syllable units is evaluated for speech recognition. Furthermore, an influence of tonal information is analyzed and the system has achieved 15% of character error rate. The authors state that, the proposed approach can reduce complexity and can increase the flexibility of a system.

A Bangla speech recognition system was developed by Anup Kumar Paul et al. (2009). Self Organizing Map (SOM) is applied to maintain the fixed length LPC trajectory and MLP is tested with 3, 4 and 5 hidden layers. For a clear
understanding, the authors have performed a comparison among different structures of neural networks and its possible solutions.

Urmila Shrawankar and Vilas Thakare (2010) have done an extensive performance comparison of feature extraction techniques used for speech recognition. The authors have discussed the merits and demerits of various feature extraction techniques such as MFCC, PLP, LPC, Linear Discrimination Analyzes (LDA) and Principle Component Analyzes (PCA). The authors state that, many researchers have proposed hybrid features to improve the recognition performance. The findings indicate that, the accuracy of the speech recognition system could be increased by using the combination of features instead of using a single feature. They strongly suggest that, more new hybrid features need to be developed for achieving better performance in robust speech recognition.

Anusuya, M. A and Katti, S. K (2012) have implemented a speaker independent Kannada speech recognition using vector quantization. Silence removal is applied by using statistical method and then Vector Quantization (VQ) technique is used to identify speech patterns. Experiments are done with VQ1 (Lawrence, R. Rabiner and Juang, B.H, 1993) and VQ2 (Lipeika, A and Lipeikiene, J, 1995) techniques for both speaker dependent and independent users. VQ1 represents a binary splitting algorithm which is used to split each cluster into two clusters, and VQ2 is done by splitting a cluster with largest average distortions into two clusters. The Euclidean distance measure is implemented to find the distance between test and reference templates. Codebook size of 32, 64 and 128 are used for evaluating both VQ1 and VQ2 techniques. Error rate has been decreased from 2.59% to 1.56% for VQ1 and 2.5% to 1.45% for VQ2 algorithm by adapting silence removal.

Vijai Bhaskar, P.et al. (2012) have implemented a speech recognition system for Telugu language using HTK. The system is trained with continuous Telugu speech, collected from different male speakers. Based on the outcomes, the system was found to be sensitive to changes in the spoken methods and environment, so improving the accuracy of the system is a challenging task. The
authors say that, various speech enhancement and noise reduction techniques should be applied for making the system more efficient, accurate and fast. The work will be extended to recognize emotional based continuous speech recognition in future.

An efficient speech recognition system for speaker-independent isolated digits was proposed by Santosh, V. Chapaneri and Deepak, J. Jayaswal (2013). The authors have computed local and global features using the Improved Features for Dynamic Time Warping (IFDTW) algorithm along with the 13 weighted MFCC coefficients. The time complexity of the recognition system has been reduced by enforcing a time-scale modification using a Synchronized Overlap-and-Add (SOLA) based technique and IFDTW. The experiments are done with TI-Digits corpus and the system has achieved the highest recognition accuracy of 99.16%. The authors say that, the proposed system is about 22 times faster than conventional techniques.

For easy understanding, the significant research works carried out in ASR for English and other Similar Languages (Except Tamil Language) are tabulated in Table 2.2.

**TABLE 2.2**

<table>
<thead>
<tr>
<th>Author(s) and Year</th>
<th>Title of the Research Work</th>
<th>Language and Dataset used</th>
<th>Feature Extraction (FE) and Speech Recognition (SR) Techniques used</th>
<th>Recognition Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ankit Kumar, Mohit Dua and Tripti Choudhary (2014)</td>
<td>Continuous Hindi Speech Recognition using Monophone based Acoustic Modeling</td>
<td>Hindi Continuous Speech</td>
<td>FE-MFCC and PLP SR-HMM</td>
<td>MFCC–95.08% PLP-85.25%</td>
</tr>
<tr>
<td>Purnima Pandit and Shardav Bhatt (2014)</td>
<td>Automatic Speech Recognition of Gujarati digits using Dynamic Time Warping</td>
<td>Gujarati Digits</td>
<td>FE- MFCC SR- DTW</td>
<td>84.44% (95.56% is achieved by improving)</td>
</tr>
<tr>
<td>Author(s) and Year</td>
<td>Title of the Research Work</td>
<td>Language and Dataset used</td>
<td>Feature Extraction(FE) and Speech Recognition(SR) Techniques used</td>
<td>Recognition Accuracy</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------------------------</td>
<td>---------------------------</td>
<td>---------------------------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Ika Novita Dewi, Fahri Firdausillah and Catur Supriyanto (2013)</td>
<td>SPHINX-4 Indonesian isolated digit speech Recognition</td>
<td>Indonesian Isolated Digit</td>
<td>FE-MFCC SR-HMM</td>
<td>50% (an Average)</td>
</tr>
<tr>
<td>Santosh V. Chapaneri (2012)</td>
<td>Spoken Digits Recognition using Weighted MFCC and Improved Features for DTW</td>
<td>English TI Digits</td>
<td>FE - Weighted MFCC SR – DTW</td>
<td>98.13%</td>
</tr>
<tr>
<td>Kuldeep Kumar and Aggarwal R. K (2011)</td>
<td>Hindi Speech Recognition System Using HTK</td>
<td>Hindi Isolated Words</td>
<td>FE-MFCC SR-HMM</td>
<td>94.63%</td>
</tr>
<tr>
<td>Sumit Kumar Ghanty, Soharab Hossain Shaikh and</td>
<td>Recognizing isolated spoken Bengali numerals</td>
<td>Bengali Isolated Numerals</td>
<td>FE – MFCC SR - DTW with Euclidean distance</td>
<td>More than 90% for speaker dependent users</td>
</tr>
<tr>
<td>Author(s) and Year</td>
<td>Title of the Research Work</td>
<td>Language and Dataset used</td>
<td>Feature Extraction (FE) and Speech Recognition (SR) Techniques used</td>
<td>Recognition Accuracy</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------------------------</td>
<td>---------------------------</td>
<td>------------------------------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Sukumar, A. A., Shah, A. F and Anto, P. B (2010)</td>
<td>Automatic Speech Recognition for Bangia Digits</td>
<td>Bangia Small Vocabulary Speaker Independent Isolated Digit</td>
<td>FE – MFCC, SR – HMM</td>
<td>more than 95% for digits(0-5) and less than 90% for digits (6-9)</td>
</tr>
<tr>
<td>Ghulam Muhammad, Yousef, A. Alotaibi and Mohammad Nurul Huda (2009)</td>
<td>Speech Recognition of Isolated Malayalam Words using Wavelet Features and Artificial Neural Network</td>
<td>Malayalam Small vocabulary Speaker independent Isolated word</td>
<td>FE- Wavelets, SR – NN</td>
<td>89%</td>
</tr>
<tr>
<td>Natasha Singh-Miller, Michael Collins, and Timothy, J. Hazen (2007)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Author(s) and Year</td>
<td>Title of the Research Work</td>
<td>Language and Dataset used</td>
<td>Feature Extraction (FE) and Speech Recognition (SR) Techniques used</td>
<td>Recognition Accuracy</td>
</tr>
<tr>
<td>--------------------</td>
<td>----------------------------</td>
<td>--------------------------</td>
<td>---------------------------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Corneliu Octavian Dumitru, Inge Gavat (2006)</td>
<td>A Comparative Study of Feature Extraction Methods Applied to Continuous Speech Recognition in Romanian Language</td>
<td>Romanian Large vocabulary Speaker independent Continuous speech</td>
<td>FE- PLP, MFCC, LPC SR – HMM</td>
<td>MFCC-90.41%, LPC-63.55%, and PLP 75.78%</td>
</tr>
</tbody>
</table>

2.3 Multilingual Speech Recognition System

The above Table showsthe research findings in ASR for different individual languages. As there is a huge demand for an ASR system in many languages, some attempts have been made to combine the language models to support multilingual speech recognition. David Imseng (2013)says that, developing robust and multilingual ASR systems is a really challenging job for current state-of-the-art ASR systems. Some important research works carried out for Multilingual Speech Recognition System is presented below.
Speech recognition for 7 languages has been developed by Ulla Uebl (1999). Experiments are done with German (G1 and G2), Italian, Slovak, Slovenian, Czech, Japanese and English while German is considered twice. Experiments for the G1 are performed using data from SPEEDATA project, which consists of dialect and non-native speakers, whereas G2 experiments are done with VERBMOBIL project dataset which involves native German speech.

Mohit Agarwal et al. (2010) have proposed a new kind of Subspace Gaussian Mixture Model (SGMM), with the parameter space constrained to a subspace of total parameter space. Experiments were done with Spanish, German and English with limited amount of training data for each language (about 10 hours). Improvements are achieved by jointly training the shared parameters of model on all languages. Word Error Rate (WER) reduction of 10.9% has been achieved using shared parameters. The authors suggest that SGMM approaches can support language independence as the parameters learned in resourceful languages. The suggested approach can be successfully reused to improve the performance for a language with limited resources.

Herve Bourlard et al. (2011) have discussed the current trends in multilingual speech processing. Authors from Idiap Research Institute say that, international language barriers should be removed for developing technology which can support government and industry. The authors also state that, Speech-To-Speech (STS) and Speech-To-Text (STT) translation are emerging key technologies for multilingual speech processing. Initially, preliminary work is conducted by modeling phonotactic constraint system. Subsequently, Out-of-Language (OOL) detection approach using confidence measures similar to OOV word detection has been implemented. Finally, MLP based language identification system has been developed.

Zoltan Tuske et al. (2014) have investigated the application of hierarchical Multi-resolution RelAtive SpecTrA (MRASTA) Bottle Neck (BN) features for under-resourced languages within Intelligence Advanced Research Projects Activity...
(IARPA) Babel project. Multilingual training on MLP-BN features on five languages such as Cantonese, Pashto, Tagalog, Turkish and Vietnamese are implemented. Experimental results prove that, a single feature stream is more beneficial to all languages than the unilingual features. Multilingual BN features improve the performance by 3 to 5% and Keyword Search (KWS) by 3 to 10% in the case of balanced corpus size. The authors have also investigated the pre-trained BN features for cross-lingual and multilingual acoustic models. BN features performed well similar to unilingual features and results proved that the simple fine-tuning step is enough to achieve comparable KWS and ASR performance on new language.

2.4 Speech Recognition using Hybrid Techniques

At present, most of the work in ASR is focusing on using a combination of techniques for achieving improvements in the outcomes. The hybrid models can take the advantage of each technique and therefore show good improvement in the recognition performance. Some of the related works carried out in ASR using hybrid techniques are tabulated in Table 2.3.

<table>
<thead>
<tr>
<th>Author(s) and Year</th>
<th>Title of the Research Work</th>
<th>Language and Dataset used</th>
<th>Feature Extraction (FE) and Speech Recognition (SR) Techniques used</th>
<th>Recognition Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asm SAYEM (2014)</td>
<td>Speech Analyzes for Alphabets in Bangla Language: Automatic Speech Recognition</td>
<td>Bangla Spoken letters</td>
<td>FE-MFCC SR- DTW+KNN</td>
<td>Speaker Dependent Users – DTW- 78% , DTW+KNN - 86% Speaker Independent Users DTW - 60% DTW+KNN - 75%</td>
</tr>
<tr>
<td>Author(s) and Year</td>
<td>Title of the Research Work</td>
<td>Language and Dataset used</td>
<td>Feature Extraction (FE) and Speech Recognition (SR) Techniques used</td>
<td>Recognition Accuracy</td>
</tr>
<tr>
<td>--------------------</td>
<td>----------------------------</td>
<td>--------------------------</td>
<td>---------------------------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Pallabi Talukdar, Mousmita Sarma and Kandarpa Kumar Sarma (2013)</td>
<td>Recognition of Assamese Spoken Words using a Hybrid Neural Framework and Clustering Aided Apriori Knowledge</td>
<td>Assamese Phoneme</td>
<td>FE – LPC SR - SOM Probabilistic Neural Network (PNN) and Learning Vector Quantization (LVQ)</td>
<td>90%</td>
</tr>
<tr>
<td>Zhao Lishuang, and Han Zhiyan (2010)</td>
<td>Speech Recognition System Based on Integrating Feature and HMM</td>
<td>Chinese Large vocabulary Speaker independent Vowels</td>
<td>FE – MFCC SR - Genetic Algorithm + HMM</td>
<td>Effective, high speed and accurate</td>
</tr>
<tr>
<td>Uma Maheswari, N., Kabilan, A. P and Venkatesh, R (2010)</td>
<td>A Hybrid model of Neural Network Approach for Speaker independent Word Recognition</td>
<td>Indian English Isolated word (50 words spoken by 20 males and 20 females)</td>
<td>FE- LPC SR- Radial Basis Function (RBF) and Brute Force algorithm</td>
<td>91%</td>
</tr>
</tbody>
</table>

2.5 Speech Recognition for Tamil Language

There have been many literatures in ASR systems for diverse languages in the globe. Unfortunately, only few works have been carried out in Tamil language when compared to English language. In order to develop a Tamil ASR, some literature about Tamil language and the main difference between Tamil and English
speech recognition should be studied. The following section discusses about Tamil language and the research works carried out, particularly for Tamil ASR.

### 2.5.1 Tamil Language

Tamil is one of the prominent and primary languages of India. It is one of the 23 nationally recognized languages in the Constitution of India. It has official status in Tamil Nadu, Sri Lanka, Malaysia and Singapore. Tamil language is structurally and phonetically different from others. Tamil vowels are classified into short, long (five of each type) and two diphthongs. Consonants are divided into three types with six in each category (hard, soft and medium). Vowel and consonants classification are performed according to the place of articulation. Totally, there are 12 vowels and 18 consonants which are combined to form 216 compound characters. Placing dependent vowel markers on either one side or both sides of the consonant forms the compound characters. There is one more special letter ‘aytham’ used in classical Tamil and rarely found in modern Tamil. Totally, 247 letters are present in Tamil alphabet (Srinivasan, A, 2009a).

### 2.5.2 Difference between Tamil and English Speech Recognition

The approach for recognizing Tamil speech is different from recognizing English speech. Speech recognition techniques developed for English language recognizes words with the aid of dictionaries. They do not attempt to recognize letters in a word, since pronunciation of an English word is not a component of the pronunciation of letters those form words. But, Indian languages are different, particularly Tamil language, where the pronunciation of words is almost composed of the pronunciation of letters. Hence, Tamil speech processing does not require a dictionary.

Accuracy of Tamil Speech Recognition can be high when compared to English speech recognition, because, in Tamil language each character has a different pronunciation, therefore, each word has a distinct pronunciation. Arun Thilak, R and Madharacim, M (2004) state that some literacy in Tamil language can
also raise the accuracy as Tamil language does not change from area to region or from person to person to a big extent.

Since the **main objective of the thesis is to develop an ASR for Tamil language**, the following section presents the research works has been carried out for Tamil speech recognition.

Henry Charles, A. P and Devaraj, G (2004) have developed Alaigal-A Tamil Speech Recognition. The research work has been proposed to produce enhanced continuous speech recognition and to support speaker-independent and device-independent capability. Spoken sentences are represented as a succession of independent acoustic phonetic units. The system recognizes spoken queries in the context of many web based applications.

Saraswathi, S and Geetha, T. V (2004) have introduced language models for Tamil speech recognition system. The use of language models in various phases of Tamil ASR namely segmentation, recognition, syllable and word level error correction phase is described. Speech signals were segmented at the phonetic level based on their acoustic characteristics. Articulatory feature based language models are applied to discover and correct the wrongly identified segmentation points. Inter and intra word based language models are used to reduce ambiguities in the recognized phonemes. These recognized phonemes were grouped together to form syllables and then words. The performance of Tamil ASR was improved moderately by using language models at different phases of speech recognition.

Lakshmi,A and Murthy,A. Hema,(2006) have presented a novel technique for building a syllable based continuous speech recognizer for Tamil language. Two different algorithms are used to segment the speech into comparable syllable like units. In order to extract accurate syllable units of speech data, a group delay based two level segmentation algorithm is proposed. Later, a rule based text segmentation algorithm is employed to automatically annotate the text corresponding to syllable units. Syllable models for isolated speech are built using
Multiple Frame Size (MFS) and Multiple Frame Rate (MFR) for all unique syllables. Better results are obtained when compared with manually segmented training data. The authors also suggest that the system development cost can be reduced by using minimum manual effort if the sentence level transcription of the speech data is available.

Likewise, Chandrasekar, M and Ponnavaikko, M (2007) have developed a technique for Spoken Tamil Character Recognition. A three layered back propagation neural network approach is used along with acoustic features of individual letters. The authors have also proposed a method, for Tamil Speech processing in three stages without using dictionaries (Chandrasekar, M and Ponnavaikko, M, 2008). It is segmentation of sentences from speech, words from segmented sentences and characters from segmented words.

Rathinavelu, A, et al. (2007) have presented a speech recognition model for Tamil language using Feed Forward Neural Networks (FFNN) with back propagation algorithm. They proposed two models, one is for giving training to the neural network and the other is for giving visual feedback. Experiments are done with 20 Tamil phonemes collected from ten children (5 boys and 5 girls) within the age group of 4-7 years. The system includes Visual Feedback module to respond to the children who are using the proposed ASR model.

Pushpa, N, et al. (2014) have developed a speech processing model for Tamil Language using Back Propagation Neural Network by adopting Semi Supervised Training. Execution of the above work involves pre-processing, using four types of filters, namely, pre-emphasis, median, average and Butterworth band-stop filter for background noise removal. The performances of these filters are measured, based on Mean Square Error (MSE) and Peak Signal to Noise Ratio (PSNR) values. LPCC feature vectors are given as an input to the FFNN for recognition of Tamil spoken word. Deep Neural Network (DNN) is proposed for LVCSR in low resource settings. The results with the specified parameters were found to be satisfactory considering less number of training data.
The other significant research works carried out for Tamil language are presented in Table 2.4.

**TABLE 2.4**

*Related Works carried out in Tamil Speech Recognition*

<table>
<thead>
<tr>
<th>Author(s) and Year</th>
<th>Title of the Research Work</th>
<th>Dataset used</th>
<th>Feature Extraction (FE) and Speech Recognition (SR) Technique used</th>
<th>Recognition Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Karpagavalli, S (2012a)</td>
<td>Isolated Tamil speech recognition using linear predictive coding and neural networks</td>
<td>10 words 20 speakers 5 times</td>
<td>FE – LPC SR - BPN</td>
<td>93.6% (100% for some words)</td>
</tr>
<tr>
<td>Karpagavalli, S (2012b)</td>
<td>Isolated Tamil Digit Recognition using Template-Based and HMM-Based Approaches</td>
<td>10 words 20 speakers 20 times</td>
<td>FE – MFCC SR - DTW HMM</td>
<td>DTW-87.8% HMM-92%</td>
</tr>
<tr>
<td>Rojathai, S (2012)</td>
<td>A Novel Speech Recognition System for Tamil Word Recognition based on MFCC and FFBNN</td>
<td>10 words 10 speakers</td>
<td>FE –MFCC SR - FFBPN</td>
<td>90%</td>
</tr>
<tr>
<td>Sigappi, A.N, and Palanivel, S (2012)</td>
<td>Spoken Word Recognition Strategy for Tamil Language</td>
<td>100 train station names in Tamil Nadu (age group-21 to 60), 20 speakers</td>
<td>FE – MFCC SR – NN HMM</td>
<td>NN-90% HMM-95%</td>
</tr>
<tr>
<td>Author(s) and Year</td>
<td>Title of the Research Work</td>
<td>Dataset used</td>
<td>Feature Extraction (FE) and Speech Recognition (SR) Technique used</td>
<td>Recognition Accuracy</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------------------</td>
<td>--------------</td>
<td>---------------------------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Dharun, V. S, and Karnan, M (2012)</td>
<td>Voice and Speech Recognition for Tamil Words and Numerals</td>
<td>Tamil Words and Numerals</td>
<td>FE – MFCC SR - DTW</td>
<td>-</td>
</tr>
<tr>
<td>Saraswathi, S and Geetha, T. V (2004)</td>
<td>Implementation of Tamil Speech Recognition System using Neural Networks</td>
<td>200 basic words 10 speaker 10 times</td>
<td>FE – MFCC SR - BPN</td>
<td>90%</td>
</tr>
</tbody>
</table>
Nevertheless, most of these works are primarily focused on the speech recognition task using MFCC and LPC features with HMM and BPN techniques. It is obvious from the above literature that, the recent attempts on auditory features like GFCC and other machine learning techniques such as KNN, SVM and random forest are not considered for Tamil speech recognition research. Also, there is a lack of literature in Tamil speech recognition for addressing the issue of noise distortions. Thus, in that respect, there is a vast demand for Tamil speech recognition to be developed under noisy environments with recent techniques.

### 2.6 Related Works on Noisy Speech Recognition

Speech signal is normally polluted by different kinds of background noise. These noises are being generated by room acoustics and human activities. Figure 2.1 shows the model of noisy speech recognition system. Occurrence of background noise in speech significantly decreases the intelligibility and quality of a signal. Also, it increases the complexity of an ASR performance.

![Figure 2.1: Model of Noisy Speech Recognition System](image)

The noisy speech recognition system is represented in Equation (2.1).

\[ y(t) = (s(t) + n(t)) * h(t) \]  \hspace{1cm} (2.1)

$s(t)$ is the original clean speech, $n(t)$ is the additive or background noise, $h(t)$ is the channel distortion and $y(t)$ is the degraded or a corrupted speech signal. Some important research works carried out for addressing noisy ASR is discussed below.

**Feature Compensation**
Xiaodong Cui and Abeer Alwan, (2005) have proposed a Feature Compensation (FC) algorithm based on polynomial regression of utterance SNR for noise robust ASR. These polynomials are used to represent the bias between clean and noisy speech features. Bias is estimated automatically from new data using an Expectation Maximization (EM) and Maximum Likelihood (ML) criterion. Initially, speech characteristics are extracted from clean data and decoded using HMM. Then, noisy speech features are compensated by regression polynomials. Comparative experiments are done with Aurora 2 (English) and German part of Aurora 3 databases. For Aurora 2 experiments, on an average a WER reduction of 16.7% and 16.5% is achieved. For Aurora 3, 15.9%, 3% and 14.6% is achieved for well matched, medium mismatched and high mismatched conditions respectively.

**Impulse Noise Reduction**

Mital. A. Gandhi et al. (2005) have presented a filtering method in time domain for detection and cancellation of impulsive noise in speech. Auto Regressive (AR) model by means of Huber M-estimator and iterative EM algorithm is used for noise detection. This method is computationally less complex than the traditional methods.

Likewise, an impulse noise removal method was proposed using soft decision and recursion algorithm by Sina Zahedpour et al. (2009). In this method, the location and amplitude of an impulse is applied by an adaptive threshold and soft decision. The position and amplitude of an impulse noise is estimated first and then an approximation of the original signal is obtained using an iterative adaptive process.

Nongpiur, R. C (2008) proposed a novel method to remove impulsive type disturbances from speech signals using multi resolution properties of wavelet transform. The wavelet coefficients correspond to impulse noise is identified and removed based on two features. They are slow time-varying nature and Lipschitz regularity of speech components. Outcomes show that, the suggested method is easily suited for removing impulsive noise from the speech signal.
Noise Adaptive Training (NAT)

Ozlem Kalinli et al. (2010) have developed a Noise Adaptive Training algorithm that can be enforced to all training data which normalizes environmental distortion as part of model preparation. NAT estimates the pseudo clean model parameters directly without depending on the clean speech features as an intermediate step. These parameters are later used with Vector Taylor Series (VTS) model adaptation for decoding noisy utterances. Experiments with Aurora 2 and Aurora 3 datasets proved that the proposed NAT method obtains relative improvements of 18.83% and 32.02% respectively.

Gammatone Filter-Bank

Hui Yin et al. (2011) have proposed a variety of features based on Gammatone filter bank. The phase modulation is represented by the sub-band Instantaneous Frequency (IF) and it is explicitly used by concatenating envelope based and IF based features. Experiments are done with Chinese mandarin digits corpus under both clean and multi-condition environments using HMM. Results proved that the proposed envelope and phase features can improve recognition rates in both conditions compared to MFCC based recognizer. Further research work focuses on phoneme recognition and large vocabulary ASR.

Noise Robustness Techniques

Tuomas Virtanen et al. (2012) have provided an extensive report on techniques used for noise robustness in Automatic Speech Recognition. The authors have presented a review of all the most important noise robust ASR approaches listed below:-

- Extraction of speech from mixture signals,
- Microphone arrays,
- Feature enhancement,
- Features based on auditory physiology and perception,
- Feature compensation,
- Model enhancement,
Factorial models for noise robust speech recognition,
Acoustic model training for robust speech recognition,
Missing-data techniques: recognition with incomplete spectrograms,
Missing-data techniques: feature reconstruction,
Computational auditory scene analysis and automatic speech recognition,
and
Uncertainty decoding.

The report has also addressed the issues involved in noise robustness and the reason for signal degradation.

Speech Recognition in natural background noise has been discussed by Meyer, Jet al. (2013). The authors have analyzed the speech intelligibility loss in spoken word lists with increasing listener-to-speaker distance in the low-level natural background noise environments. The study concentrates on some of the most essential environmental constraints that affect the spoken communication. The native French participants are involved to recognize French monosyllabic words that are spoken at 65.3 dB (at distances between 11 to 33 meters). Those words corresponded to the SNRs most revealing of the progressive effect of the selected natural noise (−8.8 dB to −18.4 dB). The experimental results have proved that the identity of vowels is mostly preserved in the above specified noisy conditions.

**Interaction with a Robot in Noisy Environment**

Mirjam Sepesy MAUCECet al. (2013) have developed speech recognition for interaction with a robot in noisy environment. Two techniques were proposed to enhance the robustness of continuous speech recognition in noisy environment. The first method is based on applying better weighting factor for the language model in the decision process, so that the accuracy of recognition can be improved. The second proposed method is based on language model adaptation. The experiments results have proved that both proposed techniques have improved the recognition accuracy by approximately 2%.
**Joint Noise Adaptive Training (JNAT)**

Arun Narayanan and DeLiang Wang, (2014) have presented a Joint Noise Adaptive Training for robust ASR. Time frequency masking is used for providing a smooth estimation of speech and noise, which is then passed as an additional feature to a DNN based acoustic model. Proposed system has improved the performance for the Aurora 4 dataset by 10.5% compared to the previous best published results. Additionally, a unified DNN framework is developed for improving acoustic modeling. The final system has outperformed the previous best results with CHiME-2 corpus by yielding 22.1% improvement.

**Modern Noise Robust Techniques**

Jinyu Li *et al.* (2014) have discussed the modern noise robust techniques for ASR using five different criteria:

1) Feature domain vs. model domain processing,
2) Use of prior knowledge about acoustic environment distortion,
3) Use of explicit environment-distortion models,
4) Deterministic vs. uncertainty processing, and
5) Use of acoustic models trained jointly with the same feature enhancement or model adaptation process for testing.

Through this study, the authors have provided an awareness of performance complexity involved in different algorithms. Also, the current challenges and future research directions in these fields are also carefully analyzed. The pros and cons of using different noise robust ASR techniques in practical application scenarios are provided as a guide for leading researchers.

The above research works were carried out using model adaptation techniques for noisy speech recognition. The subsequent section discusses about the speech signal enhancement techniques used as a pre-processor for noisy speech recognition.
2.7 Speech Signal Enhancement Techniques for Noisy Speech Recognition

In real time voice communication applications, the presence of background noise significantly reduces the quality and intelligibility of the signal. The enhancement of noisy speech signals is absolutely necessary for improving their perceptual quality. Therefore, speech signal enhancement technique should be applied as a front-end processor for noisy speech recognition to achieve better performance. Following are the important research findings on speech signal enhancement techniques proposed by various authors.

Spectral Subtraction

Maher, A. G et al. (1992) have performed a comparison of noise reduction techniques for speech recognition in telecommunications environments. The performance of telecommunications applications are significantly degraded by the environment and the telephone channel noise. The authors describe a study of various noise reduction techniques which can be applied as an input to the standard speech recognizerstrained on noise-free speech. The implementation details of speech and pitch detection that are required to support the noise reduction algorithmsare also discussed. Experimental results show that the spectral subtraction technique can produce up to 10 dB improvement in SNR, compared with 6 dB improvement provided by an adaptive line enhancement method.

Deepa, D and Shanmugam, A (2010) have implemented a speech enhancement using spectral subtraction and Partial Differential Equation (PDE). The proposed method was found to be flexible and significantly reduced the artifacts in the enhanced speech when compared with conventional Power Spectral Subtraction (PSS) method. The resultant signal has improved both speech quality and intelligibility. The algorithm have produced less computational complexity and can able to adapt for non-stationary noise and Multi Band Spectral Subtraction (MBSS) method. The authors state that, the developed method can be highly useful
for digital hearing aids, because the hearing impaired listeners require 5 dB to 10 dB higher SNR than normal hearing people.

**Adaptive Filtering**

Jagan Naveen, *et al.* (2010) have performed noise suppression in speech signals using Least Mean Squares (LMS) and Recursive Least Squares (RLS) algorithms. These two algorithms are implemented and compared based on SNR and tap weights of FIR filter. Experimental results show that, RLS algorithm produces highest SNR and it outperforms the LMS algorithm. In contrast, LMS has offered faster convergence than RLS algorithm. The authors state that, the optimum Mu and Lamda values can be obtained by adjusting the FIR Tap weights. In this work, the optimum Mu values for LMS were found to be 0.9 for FIR Tap-7 and 0.01 for FIR Tap-3. Likewise, the optimum Lamda values for RLS were found to be 0.92 for FIR Tap-3 and 0.94 for FIR Tap-7.

Sayed, A. Hadei and Lotfizad, M (2010) have implemented a family of adaptive filter algorithms for speech enhancement. The authors say that, as the signal characteristics changes rapidly, the utilization of adaptive algorithms is well suitable, since they converge very fast. During recent years, the trade-off between computational complexity and the convergence speed in adaptive filtering has been addressed. Accordingly, the authors have developed a new approach for noise cancellation by implementing two novel adaptive filtering algorithms, namely, Fast Affine Projection (FAP) and Fast Euclidean Direction Search (FEDS) algorithms. The proposed techniques are compared with the classical adaptive filters, such as LMS, NLMS, Affine Projection (AP) and RLS algorithms based on the time evolution of filter taps and MSE. The experimental results have proved that the developed techniques are comparable with the existing adaptive algorithms.

**Combinational Adaptive Filtering**

Recently, Raghavaraju, A and Bhavani Thota (2014) have implemented a Speech Enhancement using Combinational Adaptive Filtering Techniques. The
authors have used various adaptive filtering techniques individually as well as in combination for speech noise cancellation. The combinations of algorithms are as follows: Least Mean Square (LMS), Normalized LMS (NLMS), Variable Step Size LMS (VSLMS) and Variable Step Size Normalized LMS (VSNLMS). The capability of the above algorithms are measured based on SNR improvement. The experimental results show that the NLMS algorithm was found to be better than LMS based algorithm in noise reduction.

Kalman Filtering

Sharon Gannot et al. (1998) have implemented iterative and sequential kalman filter-based speech enhancement algorithms. Some extensions, modifications, and improvements of previous work based on Kalman filtering were discussed. Initially, estimate-maximize method is applied to estimate the spectral parameters of the speech and noise parameters iteratively. Next, a sequential based computationally efficient, gradient descent algorithm is implemented. Extensive comparative analyzes of the iterative and sequential algorithms with the existing techniques were discussed.

Enhancing the Reverberant Speech

Yegnanarayana, B et al. (2002) have developed a technique for enhancing the reverberant speech using Linear Prediction (LP) residual signal. The authors have proposed an approach for enhancing the speech corrupted by noise and reverberation from multiple microphones. The developed technique works based on exploiting the features of the excitation source in speech production. The characteristics of voiced speech can help to derive logical information from the LP residuals of the degraded speech information acquired from different microphones. In this work, a weight function is derived from the coherently added signal, for which the time-delay between a pair of microphones is calculated using the data provided by LP residual. Finally, the enhanced speech signal is generated by exciting the time varying all-pole filter with the weighted LP residual.

Single Channel Noise Reduction Algorithm
Yi Hu and Philipos, C. Loizou (2007b), have performed a comparative intelligibility study of single-microphone noise reduction algorithms. The IEEE sentences and consonants were corrupted by babble, car, street and train noise at 0 dB and 5 dB SNR. The spectral subtractive, sub-space model, statistical model based and Wiener-type algorithms were involved for performing speech signal enhancement. Based on the outcome, the authors found that there were no significant improvements in speech intelligibility, by using the above mentioned algorithms. Based on the overall analyzes and consonant confusion matrices, the authors state that, the noise reduction algorithms need to improve the place and manner feature scores in order to enhance the performance of speech intelligibility.

Junfeng Li et al. (2011) have done a comparative intelligibility investigation of single channel noise reduction algorithms for Chinese, Japanese and English datasets. Given the different perceptual cues used by native listeners of different languages including tonal languages, it is of interest to examine whether there are any language effects when the same noise reduction algorithm is used to process noisy speech in different languages. Clean speech signals were first corrupted by three types of noise at two SNR and then processed by five single channel noise reduction algorithms. The processed signals were finally presented to normal-hearing listeners for recognition. Speech intelligibility assessment showed that most of the noise reduction algorithms did not improve speech intelligibility. Previous experiments with English language using wiener filtering have offered limited but statistically significant improvements in intelligibility for car and white noise conditions.

**Dual-Microphone Speech Enhancement**

A novel dual-microphone speech enhancement technique is proposed by Nima Yousefian and Philipos, C. Loizou (2012). The coherence between the target and noisy signals were used as a measure for noise reduction and applied to closely spaced microphones. The developed algorithm was found to be simple to implement and has the applicability of handling multiple interference, also it does not requires the statistics on noise estimation.
The proposed algorithm was evaluated using intelligibility listening tests and evaluated against a most popular beam forming algorithm by involving Normal Hearing (NH) listeners. The proposed algorithm was found to be significantly better in producing high intelligibility, when compared with beam forming algorithm, mainly when the signal is corrupted by multiple noise sources or competing talkers. The experimental outcomes based on objective and subjective measures indicate that, the proposed coherence-based algorithm performs better for both hearing impaired and cochlear implant devices.

Performance comparisons between various speech enhancement algorithms are proposed over the years, but it is very difficult to predict which technique is best due to following factors:

- Different types of noise,
- SNR dB levels,
- Methodology, and
- Database.

To facilitate the above constraints, Yi Hu and Philipos, C. Loizou et al. (2007) have developed a noisy speech corpus, which is suitable for the evaluation of speech enhancement algorithms. This corpus is subsequently used for evaluating 13 speech enhancement methods encompassing four classes of algorithms, namely

- Spectral subtractive,
- Subspace,
- Statistical model based algorithm, and
- Wiener type algorithm.

The subjective evaluation was performed using ITU-T P.835 methodology designed to assess the speech quality in three different dimensions, namely, signal distortion, noise distortion and overall quality.

**Wavelet Transform**
Slavy, G. Mihovet al. (2009) performed a de-noising of noisy speech signals by using Wavelet Transform. The use of wavelet transform in de-noising the speech signals contaminated with common noises is investigated. The authors state that, the wavelet-based de-noising with either hard or soft thresholding was found to be the most effective technique for many practical problems. The experimental results with a large database of reference speech signals contaminated with various noises in several SNR dB conditions are presented. The authors also insist that, the power spectrum estimation using a wavelet-based de-noising may be applied as an important approach for better speech signal enhancement. The research work will be extended to be applied for the practical research on speech signal enhancement for hearing-aid devices.

Rajeev Aggarwal et al. (2011) have implemented a Discrete Wavelet Transform (DWT) based algorithm using both hard and soft thresholding for denoising. Experimental analyzes is performed for noisy speech signals corrupted by babble noise at 0dB, 5dB, 10dB and 15dB SNR levels. Output SNR and MSE are calculated and compared using both types of thresholding methods. Experiments show that soft thresholding method was found to be better than a hard thresholding method for all the input SNR dB levels involved in the work. The hard thresholding method has extended a 21.79 dB improvement while soft thresholding has achieved a maximum of 35.16 dB improvement in output SNR.

Jai Shankar, B and Duraiswamy, K (2012), have proposed a de-noising technique based on wavelet transformation. The noise cancellation method is improved by a process of grouping closer blocks. All the significant information resides in each set of blocks are utilized and the vital features are extracted for further process. All the blocks are filtered and restored in their original positions, where the overlapping is applied for grouped blocks. The experimental results have proved that the developed technique was found to be better in terms of both SNR and signal quality. Moreover, the technique can be easily modified and used for various other audio signal processing applications.
Complex Wavelet Transform

Reshad Hosseini and Mansur Vafadust, (2008) have developed an almost perfect reconstruction filter bank for non-redundant, approximately shift-invariant, complex wavelet transforms. The proposed novel filter bank with Hilbert pairs wavelet filters do not have serious distributed bumps on the wrong side of power spectrum. The redundancy of an original signal is significantly reduced and the properties of proposed filter bank can be exploited in different signal processing applications.

Amplitude Modulation Spectral Analysis

Jorg-Hendrik Bach et al. (2010) have performed a robust speech detection in real acoustic background noise by using a sub-band Amplitude Modulation Spectral (AMS) features and trained discriminative classifiers. Performance evaluation is done for SNR ranges from -10 dB to 20 dB. The findings from the results are:

- Generalization to novel background classes with AMS features yields 10 dB improvements in SNR compared to MFCC features,
- For well-known background noise, AMS and MFCCs achieve similar performance, and
- Standard voice activity detection (ITU G729.B) has significant performance in real acoustic backgrounds.

Ideal Binary Mask (IBM)

Following research finding has been analyzed by Ning Li and Philipos, C. Loizoua, 2008). The authors have discussed the factors influencing intelligibility of ideal binary-masked speech. Specifically, the effect of local SNR threshold, input SNR level, type of masker and errors introduced in estimating the ideal mask are examined. The authors state that, the performance was mostly affected when the masker dominated Time-Frequency (T-F) units were wrongly labeled as target-dominated T-F units. Performance plateau near 100 % correct for SNR thresholds ranging from -20 dB to 5 dB. This pattern directs the listener's attention to identify
where the target speech is present and enables them to segregate speech effectively in multi talker environments.

De Liang Wang *et al.* (2009) have done two experiments on signal separation technique to assess the effects of IBM on speech intelligibility. Both Normal Hearing (NH) and Hearing Impaired (HI) listeners at different kinds of background interference are involved for the performance analyzes. The authors have proved that the IBM technique leads to substantial reductions in speech reception threshold for both NH and HI listeners. The technique has better noise reduction in a cafeteria background than for a speech shaped noise. Furthermore, listeners with hearing loss benefited more when compared with listeners with normal hearing, particularly for cafeteria noise. IBM nearly equalizes the speech intelligibility performances of NH and HI listeners in noisy backgrounds. Experimental results proved that ideal binary masking in the low frequency range yields high intelligibility improvements than in high frequency range, particularly for listeners with hearing loss. Findings from the experiments have major implications for understanding speech perception in noise, Computational Auditory Scene Analyzes (CASA), speech enhancement and hearing aid design.

**Teager Energy Operator (TEO)**

Huan Zhao *et al.* (2011) have proposed an improved speech enhancement method based on Teager Energy Operator (TEO) and Perceptual Wavelet Packet Decomposition (PWPD). The proposed technique works in the frequency domain, therefore, better results were achieved. In this method, initially a modified mask construction method is applied in order to protect the acoustic cues present in the low frequencies. Subsequently, a level-dependent parameter is introduced to fine-tune the thresholds in the noise distribution feature. The experimental outcomes have proved that the proposed method was better in improving the SNR and PESQ, and also minimizes the computation load.
An Approach for Noisy Speech Enhancement through soft thresholding function by Employing the Teager Energy Operation (TEO) on Wavelet Packet (WP) was proposed by Tahsina Farah Sanam and Celia Shahnaz (2012). The authors have used a threshold, which is statistically determined by using the TEO on the Wavelet Packet coefficients of noisy speech. The obtained threshold is further applied with WP coefficients for speech enhancement. The experiments are done with four type of noises, namely, white, car, pink, and babble. The experimental results based on standard objective measures, spectrogram representations and subjective listening tests have proved that the proposed technique yielded better performance compared to existing thresholding based speech enhancement methods for both high and low level SNR.

**Noise Reduction for Cochlear Implant**

Hua Ye *et al.* (2013) have implemented a noise reduction for Cochlear Implant (CI) recipients to achieve acceptable speech perception in noisy environments. The authors state that the algorithms based on signal representations other than T-F representations may also be able to provide comparable speech perception and listening quality improvements. A dual tree complex discrete wavelet transform is implemented, using wavelet shrinkage coefficient based on a statistical approximation of the variance attribute of noise. The suggested technique was evaluated by comparing its performance with the existing wavelet based algorithms. Speech Transmission Index (STI) of the proposed algorithm is significantly better than the existing algorithms used for comparison. An average significant improvement in speech perception of 1.9 dB was achieved for weighted noise.

Mike Brookes *et al.* (2008) have performed an extensive literature review on speech cleaning. The report provides a summary of the literature survey carried out for speech cleaning in three aspects:

- First, the current state-of-the-art methods that are used to evaluate the quality and intelligibility of the enhanced speech are discussed,
Second, the principal approaches that are proposed for speech signal enhancement are reviewed,

Third, the techniques used for estimating the characteristics of the background noise are reported, and

Finally, a brief overview of the commercial speech cleaning systems and the speech enhancement databases that are widely used for evaluating the noise cancellation techniques are presented. The speech enhancement databases involved in the study are as follows:

- NOISEX,
- TIMIT, NTIMIT and CTIMIT,
- Aurora,
- ITU-T Coded Speech Database,
- Noizeus,
- SpEAr Database, and
- RSPL Noises.

Lakshmikanth, S. et al. (2014) have presented a review on noise cancellation in speech signal processing. Findings from the review are as follows:

- RLS algorithm produces high SNR improvement,
- RLS is superior to Least Mean Squares (LMS),
- LMS has advantage of faster convergence when compared with RLS,
- Optimum Mu (LMS) and Lambda (RLS) values can be obtained by fixing the FIR tap weight, and
- Wavelet transform and Empirical Mode Decomposition (EMD) performs very well for noisy environments.

**Evolutionary Algorithms**

Kanchan Malviya, et al. (2014), has implemented a speech enhancement using adaptive filter based on evolutionary algorithms. The Gradient based adaptive algorithms like LMS, NLMS and RLS are most popularly used, because
of its simplicity in computation. These algorithms provide only one possible solution for each iteration based on the estimated error, and not suitable for multimodal error surface. The authors have applied an optimization algorithm to increase the probability of encountering the global optimum. Constant Weight Inertia Particle Swarm Optimization (CWI-PSO), Linear Decreasing Inertia PSO (LDI-PSO), Non-Linear Decreasing Inertia PSO- (NLI-PSO), Exponential PSO and Bacterial Foraging Optimization (BFO) algorithms are involved in the research work. Based on the results, it is observed that the Evolutionary approach is found to be effective and can improve the performance of adaptive filter. Based on the comparative analyzes of the above mentioned algorithms, the Exponential-PSO algorithm gives the better performance based on fidelity parameter.

Chapter Summary

In this chapter, the review of literature is carried out in six different categories. The significant research findings from various authors were presented. The strength and the limitations of various techniques were observed based on the literature. Also, the huge demand for the proposed work is identified and given below:

Observation from the Literature:

It is observed from the above literature that, there is a huge demand for the following tasks:-

- Robust techniques to support speaker and environment independent ASR,
- ASR for Tamil Language using recent feature extraction and recognition techniques under noisy environments,
- Performance evaluation of noisy ASR by involving negative SNR dB levels, and
- Efficient speech signal enhancement techniques for noisy ASR.

In order to carry out the above tasks, the research design is framed in Chapter 3, entitled Methodology.