Automatic speech recognition (ASR) based approach for speech therapy of aphasic patients: A review

Cite as: AIP Conference Proceedings **1883**, 020028 (2017); https://doi.org/10.1063/1.5002046
Published Online: 14 September 2017

Norezmi Jamal, Shahnoor Shanta, Farhanahani Mahmud, et al.

Automatic Speech Recognition (ASR) based Approach for Speech Therapy of Aphasic Patients: A Review

Norezmi Jamal[1, b)], Shahnoor Shanta[2, a)], Farhanahani Mahmud [1, c)] and MNAH Sha'abani[3, d)]

[1]*Microelectronic and Nanotechnology Shamsuddin Research Centre (MiNT-SRC), Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia*
[2]*Department of Electrical Engineering, Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia*
[3]*Department of Electrical Engineering, Centre of Diploma Studies, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat , Johor, Malaysia*

[a)]Corresponding author: shahnoor@uthm.edu.my
[b)] jamalnorezmi@gmail.com
[c)] farhanah@uthm.edu.my
[d)] nhafiz@uthm.edu.my

Abstract. This paper reviews the state-of-the-art an automatic speech recognition (ASR) based approach for speech therapy of aphasic patients. Aphasia is a condition in which the affected person suffers from speech and language disorder resulting from a stroke or brain injury. Since there is a growing body of evidence indicating the possibility of improving the symptoms at an early stage, ASR based solutions are increasingly being researched for speech and language therapy. ASR is a technology that transfers human speech into transcript text by matching with the system's library. This is particularly useful in speech rehabilitation therapies as they provide accurate, real-time evaluation for speech input from an individual with speech disorder. ASR based approaches for speech therapy recognize the speech input from the aphasic patient and provide real-time feedback response to their mistakes. However, the accuracy of ASR is dependent on many factors such as, phoneme recognition, speech continuity, speaker and environmental differences as well as our depth of knowledge on human language understanding. Hence, the review examines recent development of ASR technologies and its performance for individuals with speech and language disorders.

INTRODUCTION

Over the last few decades, automatic speech recognition (ASR) has been gained a surge of interest among inventors and researchers in the speech processing research area. This is due to the widely used to many applications. For example, ASR system has been used in telephony [1], military [2] and customer service [3]. ASR system based approach is also getting demand in rehabilitation that can help people who suffer from communication disorder especially aphasic patients to do speech therapy and cognitive exercise [4-6]. The individuals with the disorder only need to have a computer to conduct independently their speech therapy from home by creatively responding to the feedback provided by the system about the subject's incorrect word production. The motivations of this recent technology in speech pathology applications are to reduce cost and to increase articulation precision effectively [5, 7]. ASR system has the ability to process human speech signal and transforming it to the desired text message transcription efficiently and accurately.

However, ASR system still has a lack of performance in terms of its accuracy and robustness. The reason is most speech recognition systems easily perceiving to the acoustical environments [8, 9]. Even though the main challenge of ASR system is noises, the general issues like speech task, speaker mode, size of vocabulary and speaking style also influence ASR system performance [10]. Other specifics issues like speech intelligibility, severity and speech

Advances in Electrical and Electronic Engineering: From Theory to Applications
AIP Conf. Proc. 1883, 020028-1–020028-8; doi: 10.1063/1.5002046
Published by AIP Publishing. 978-0-7354-1563-8/$30.00

variability due to speech impairment also can affect the performance of ASR system as discussed in the reference [10]. On the other hand, the different techniques or approaches that have been used in speech recognition also can affect the ASR system performance [11]. Based on the aforementioned issues, the main study of this paper is to investigate the previous related research approaches of ASR system for the treatment of Aphasia. It is crucial to improve and enhance ASR system based approach for speech therapy of aphasic patients by identifying the suitable method of speech recognition and investigating the impact of aphasia on the accuracy of ASR.

This paper is organized as follows: Section 2 discusses the Aphasia and its therapy including the conventional and the modern approach. Section 3 discusses the architecture of ASR system and how it works. Section 4 presents the related research on ASR system that have so far been used in the speech therapy of aphasic patients. Lastly, Section 5 presents the concluding remarks about the potential of ASR system for aphasia including the limitations and the future directions.

APHASIA AND ITS THERAPY

Aphasia is loss of language functioning and having a problem to communicate orally resulting from a stroke or brain injury in the absence of sensory, motor, or cognitive impairments. It is estimated that from 21% to 38% of stroke patients suffers from aphasia [12]. Symptoms of aphasia are quite subjective, varying from one speaker to the other. Some people with aphasia know a word which might feel like having on tip of the tongue but it is just difficult for them to get the right words out. On some cases, in terms of speech production, some people with aphasia often exhibits the inaccurate of phonemic, misrepresentations of articulation, and speech disfluencies [6]. During the first year of post stroke event, most people with aphasia could get better improvement of their condition. For example, authors in reference [12] studied that the greatest amount of spontaneous recovery occurs in the first three months following stroke compared to the late stage of aphasia. Next, the aphasia recovery also depends on the intensity of speech therapy. The authors in reference [13] have studied the rehabilitation of aphasia effects and discovered that high intensity of aphasia therapy over a short time period has greater impact on recovery than less intensity of therapy over a longer time period.

Conventional Speech Therapy

Traditionally, people with aphasia undergo their speech therapy practice with the speech-language pathologists (SLPs) [14].This speech therapy involves face to face speech therapy, together with the manual handout linguistic task [7]. Mostly aphasic patients need to do their speech cognitive exercise in front of speech pathologist at the rehabilitation center. As a result, people with aphasia have time limits to do their speech therapy practice, which need to follow the SLP's schedule. Otherwise, the use of speech therapist service on long term basis will lead to the high cost [5]. In many of the times the conventional speech therapy might not be convincing and effective [14]. Plus, the speech therapy based on music and game are just for making aphasic patients happy, relaxed and motivated. The example of conventional speech therapies are (i) Constraint-induced Language Therapy [15], (ii) Melodic Intonation Therapy [16], (iii) Reading Treatment, (iv) Script Training [17] and Computerized treatment without ASR system [18]. The conventional computerized treatment offers a promising addition to in person therapy [18]. Unfortunately, most software programs were developed in the computer to aid the verbal exercise of aphasia are incapable to provide the type of feedback administered via ASR system by SLPs [18]. Due to this, the drawback of conventional computerized therapy are self-monitoring their verbal output and may cause aphasic patients to cultivate bad habits in their speech exercise without having interaction with SLPs [5].

Modern Speech Therapy

In modern approach to speech therapy, computer based speech therapy with automatic speech recognition (ASR) has more recently been developed in order provide intensive speech training to the individuals with aphasia. This modern technology is designed to enhance or supplement the conventional method, which involves individual therapy session with SLPs. Normally, people who suffer from aphasia need enough treatment and practice to remain fairly effective in their communication. Unfortunately, this becomes difficult for the aphasic individuals due to financial limitations, travel costs, scheduling constraints, shortage of SLPs, and health issues [5]. With ASR technology, aphasic patients have no time limitation for doing speech therapy practice from the comfort of their own home by creatively responding to the speech task designed for aphasic patients. Based on the literature review, authors in the references [4-7] have conducted several studies on ASR based approach for speech therapy of aphasic patients. The details of related ASR based approach for aphasic condition will be discussed in Section 4. In the studies conducted so far, the ASR based approach for speech therapy, speech diagnosis or speech assessment are deemed necessary for patients with aphasia. There may not be a 100% reversal results from therapy and rehabilitation, but an improvement will always be a good outcome. It is not easy living with someone having this condition, let alone having this condition ourselves, it is important to bring positive changes by giving our assistance and patience to these people in order to improve their way of life.

AUTOMATIC SPEECH RECOGNITION (ASR)

Over the past several decades, most research outcomes in the speech processing applications have emphasized the use of automatic speech recognition (ASR). There is a large volume of published studies describing the ASR architecture [19-21]. Basically, ASR system architecture consists of two main parts, which are (i) front-end process and (ii) back end process as shown in Figure 1 [22].

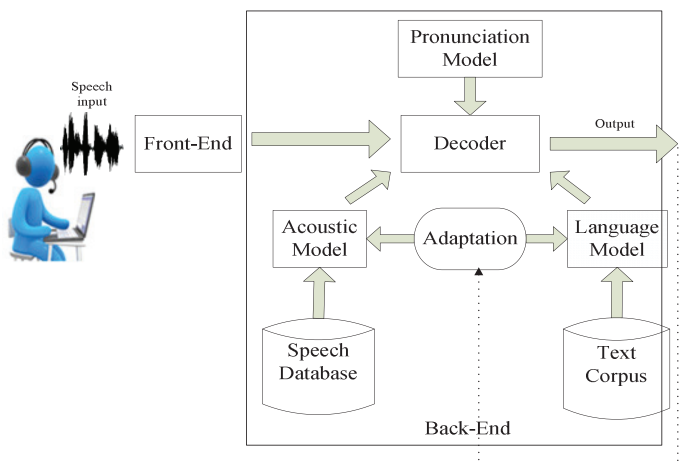

FIGURE 1. A general block diagram of ASR system [22]

Front-end Process of ASR

At the front-end process, the speech signals need to be parameterized via the pre-processing, feature extraction and feature selection techniques [22]. The speech signal normally in the form of analogue speech signal, needs to be captured and converted into digital form that can be read by the computer using microphone. Next, it is required to be sampled, framed, filtered and analysed with the sampling frequency between 8 kHz and 16 kHz [23]. The choice of sampling frequency normally depends on the use of computer and microphone specifications. Higher sampling rate will lead to higher processing rate [23]. Since the speech signal is non-stationary or time varying signal, the speech signal needs to be framed for assuming short-time stationarity. The typical value of frame size that is normally used is between 15 ms and 40 ms [8]. The choice of speech frame size should not be too short or too long. The reason is that if the frame size is too short, it would result in unreliable spectral estimation for window having insufficient samples; if the frame size is too long, it would cause speech signal having abrupt changes resulting in loss of precision [23]. So, the task of choosing the best window size and frame period is crucial for the designing of speech recognizer. It will be used for back end processing and would lead to the reliable speech recognition.

The process of transforming the pre-processed speech signal into numerical values that could be used for encoding a speech signal is known as feature extraction. There are many possible classes of features that have been reported to be used for speech recognition. Mel Frequency Cepstral Coefficient (MFCC) inspired by auditory modelling is one of the most prominent feature extraction techniques as it is found to be more efficient and simple compared to the time domain features [24, 25]. Other well-known feature extraction techniques such as Perceptual Linear Prediction (PLP), Relative Spectral Transform Perceptual Linear Prediction (RASTA PLP), Linear Predictive

Coding Coefficient (LPCC) and Discrete Wavelet Transform (DWT) have also been reported to be used in the speech recognition [11]. Power Normalized Cepstral Coefficients (PNCC) is one of the recent feature extraction techniques which has been found to rectify the noise issue [26]. On the other hand, some studies have combined the features [27] in order to maximize the recognition accuracy. Table 1 shows the summary of the feature extraction techniques presenting a comparison in terms of their advantages and disadvantages.

TABLE 1. Advantages and disadvantages of different feature extraction techniques

Techniques	Advantages	Disadvantages
MFCC	Can discriminate the repetitions and prolongations speech signal [28] Small correlation [25]	Not robust enough in noisy environments [25]
PLP	More susceptible to human hearing [29]	Resultant feature vectors are dependent, which best suited with Deep Neural Network (DNN) classification [30]
RASTA PLP	Capable to deal with various kinds of noise [31]	Not really robust [11]
LPCC	The de-correlated features components [11]	Inadequate linear scales [11]
DWT	Capable to compress a signal without major degradation [11]	Not flexible enough [11]
PNCC	Better than RASTA-PLP and MFCC in terms of its noise and reverberant cancellation [26]	Need speech enhancement algorithm techniques [26]

Selecting an optimal set of features plays a significant role at the front-end process of speech recognition for real-time applications. The purpose is to avoid inclusion of correlated or redundant features which results in reduced computation burden, decrease in training time and improved classification accuracy. There are many techniques that have been reported in the literature for feature selection or optimization. For example, Mirhassani and Ting proposed fuzzy-based discriminative feature representation to get the optimum features for children's speech [32]. Other feature selection techniques such as fisher's ratio measure and Linear Discriminant Analysis (LDA) are also widely used for feature selection in speech recognition [33, 34]. The principal component analysis (PCA) can be used to overcome the redundancy issues [35].

Back-end Process of ASR

At the back end process of ASR, language, pronunciation, and acoustic model are the most essential processes, the combined implementation of which could recognize the speech based on extracted features. Language model normally consists of grammar and linguistic properties [22]. This model is required to recognize not only the phonemes that create the input speech signal, but also to measure between the levels consisting of either trigram, words or even sentences. Thus, the modelling of a language is necessary in order to produce meaningful representations of the speech signal combined with Hidden Markov Model (HMM) or can be used as an extension. Next, pronunciation (lexicon) model is also needed to produce optimal sequence of words that compose the system's final output during recognition. Practically, the sequence of symbols generated by the acoustic component is compared with the set of words present in the lexicon. Thus, a lexicon which is also known as a dictionary is used to provide the mapping between words and phones [22]. It contains information about which words are known to the system and also how these words are pronounced.

The acoustic model mainly involves the classification of the basic speech units including phones, syllables and the acoustic observations based on the extracted features and language models [11, 22]. This classification can be categorized into two different approaches: either generative or discriminative approach. Basically, the generative approach is based on the probability distribution with the given observations and the class labels while discriminative approach need to have a conditional distribution using a parametric model [11]. The well-known methods that are based on the generative approach are Hidden Markov Model (HMM) and the Gaussian Mixture Model (GMM) [11]. On the other hand, Artificial Neural Network (ANN), Deep Neural Network (DNN) [36], Multilayer Perceptron (MLP) and Support Vector Machine (SVM) can be categorized as the discriminative approach. Recently, the hybrid models have also been proposed by many researchers to be applied in ASR [37-39]. The idea is to improve the ASR performance by combining the strengths of both the approaches [11]. The summary of the comparative study for the hybrid acoustic models is presented in Table 2.

TABLE 2. Advantages and disadvantages of different acoustic model techniques

Techniques	Advantages	Disadvantages
GMM-HMM	Only perform in clean environment with MFCC features [38]	Sensitive to noisy environment and not really robust[38]
DNN-HMM	Less word error rate [40]	Not really robust for many layer and higher cost computational [41]
MLP-HMM	Outperform in clean and noisy environment [38]	Complicated due to MLP does not has any specific rule [38]
SVM-HMM	Higher accuracy [39]	Not suitable for MFCC features [42]

ASR BASED APPROACH FOR APHASIC PATIENTS

During the past few years, a large body of knowledge has become readily available on the speech recognition and assessment of dysarthria [43-46] and stuttering [47-50]. The people suffering from speech and language disorders have high speaker to speaker variations and involves data scarcity compared to the data collected from the individuals without the conditions [51]. To deal with these issues, several methods of pattern recognition and classifications have been used in the diagnosis and prognosis of dysarthria and stuttering respectively. For example, Shahamiri, S. R., and Salim, S. S. B [45] used Artificial Neural Network (ANN) to recognise dysarthria condition with Mel Frequency Cepstral Coefficient (MFCC) features while Ai *et. al.* [50] implemented K-Nearest Neighbour (KNN) and Linear Discriminant Analysis (LDA) as speech recognisers for stuttering conditionn with MFCC and Linear Predictive Coding Coefficient (LPCC) features. On the other hand, there is only limited recent literature available on the ASR based approach for speech therapy of aphasic patients as stated earlier in Section 2. These studies involve ASR for aphasic patients conducted in different languages such as English [5, 7, 51], Portuguese [4] and Cantonese [6]. Considering the results as summarized in Table 3, it seems that the performance of ASR based approach for aphasic patients still needs some improvement in terms of its speech intelligibility and speaker adaptability.

TABLE 3. Previous works of ASR based approach for Aphasic Patients

Authors	Speech Task	Feature Extraction	Acoustic Modelling	Error Rate
Le *et. al.* [51]	Continuous	MFCC-LDA	GMM-HMM	39.7%
			DNN-HMM	42.9%
Lee *et. al.* [6]	Continuous	MFCC-LDA	GMM-HMM	58.2%
			DNN-HMM	57.8%
Abad *et. al.* [4]	Isolated	PLP-Rasta-Modulation Spectrogram (MSG)	MLP-HMM	21.0%

Le *et. al.* [51] used University of Michigan Aphasia Program (UMAP) database, which consists of large vocabulary of continuous speech in English collected from individuals with Aphasic condition (11 males, 6 females, age 58 ± 14). The audio had been recorded using the tablet's built-in microphone with a 44.1 kHz of sampling rate [51]. The authors tried to improve the ASR of Aphasic speech using DNN-HMM with MFCC and LDA features. Unfortunately, DNN-HMM did not give a promising result as shown in the summary Table 3 reinforces the data scarcity problem in aphasic speech recognition. [51]. The error rate of using DNN-HMM was 42.9% which is larger than GMM-HMM method due to the acoustic model is not really robust enough with the extracted features. There is possibility that the features had been distorted by noises during front-end process.

Next, Lee *et. al.* studied ASR based approach for speech therapy of Cantonese speaking Aphasic patients with different acoustic models [6]. They used Cantonese Aphasia Bank which consists of spontaneous oral narratives recorded file for 149 unimpaired native Cantonese speakers and 104 individuals with post-stroke aphasia based on given speech task [6]. The audio had been recorded using a head-worn condenser microphone and a digital recorder with 44.1 kHz of sampling rate [6]. They found that GMM-HMM and DNN-HMM produced 58.2% and 57.8% of error rate respectively with MFCC and LDA features. It showed that acoustic models are not the most critical issue producing the low accuracy, which only having smallest differences of error rate. Based on Table 3, the error rate of the authors in the reference [6] is higher than authors in the reference [51]. The reasons are speaking style and language models are the main challenges in getting high accuracy in ASR.

Conversely, Abad *et. al.* [4] carried out their research on automatic word naming recognition for an on-line aphasia treatment system based on acoustic keyword spotting approaches. The audio of Portuguese speakers had been collected at two different phases which are 1) the speech of 8 aphasic patients had been recorded in small room of wooden walls and 2) the speech of 8 aphasic patients had been recorded in a larger room [4]. The authors reported that the word verification rate of phase 1 is higher than phase 2. This is due to the environment of phase 2 is easily distorted by noise at broader space. It is possible to achieve high performance word verification rates for different types of patients and acoustic conditions [4]. A smaller average error rate of 21.0 % was also found when dealing with the isolated speech compared to continuous speech database. In short, simple speech task will lead to higher recognition accuracy and vice versa.

CONCLUSION

This review paper gives a brief overview of ASR based approach for speech therapy of individuals with Aphasia. In Aphasic condition, the affected individuals suffer from speech and language impairment, which can be characterized by disturbance in interpretation as well as formulation of language symbols. They need prompt and intensive course of speech therapy and practice for rehabilitation which is most of the times carried out by a registered speech-language pathologist. Due to several factors such as cost and efficient time allocation, ASR based approaches have the potential to augment the conventional method as discussed in Section 2. The general architecture of ASR as a tool for speech therapy for Aphasia based on the past studies has also been discussed in this review paper. The major constraints of ASR that have been found from this review are to achieve high accuracy and more robust speech recognition. There is still scope for further development and improvement, especially at the front-end process of the ASR. A few important elements need to be considered in order to compute more robust and optimal feature set, especially for the treatment of aphasia.

ACKNOWLEDGMENTS

Sincerely to express highly appreciation to Universiti Tun Hussein Onn Malaysia (UTHM) for funding this research work under Short Term Grant (U530) and also Microelectronic and Nanotechnology Shamsuddin Research Centre (MiNT-SRC) for providing research space.

REFERENCES

1. R. B. Garberg and M. Yudkowsky, "Method for automatic speech recognition in telephony," ed: Google Patents, 1998.
2. B. Beek, E. Neuberg, and D. Hodge, "An assessment of the technology of automatic speech recognition for military applications," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 25, pp. 310-322, 1977.

3. C. P. Gusler, A. H. I. Rick, and T. M. Waters, "Employing speech recognition and capturing customer speech to improve customer service," ed: Google Patents, 2005.
4. A. Abad, A. Pompili, A. Costa, I. Trancoso, J. Fonseca, G. Leal*, et al.*, "Automatic word naming recognition for an on-line aphasia treatment system," *Computer Speech & Language,* vol. 27, pp. 1235-1248, 2013.
5. D. Le, K. Licata, C. Persad, and E. M. Provost, "Automatic Assessment of Speech Intelligibility for Individuals With Aphasia," *IEEE/ACM Transactions on Audio, Speech, and Language Processing,* vol. 24, pp. 2187-2199, 2016.
6. T. Lee, Y. Liu, P.-W. Huang, J.-T. Chien, W. K. Lam, Y. T. Yeung*, et al.*, "Automatic speech recognition for acoustical analysis and assessment of cantonese pathological voice and speech," in *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*, 2016, pp. 6475-6479.
7. M. C. Linebarger and J. F. Romania, "Aphasia therapy system," ed: Google Patents, 2007.
8. L. R. Rabiner and R. W. Schafer, *Theory and applications of digital speech processing* vol. 64: Pearson Upper Saddle River, NJ, 2011.
9. E. Vincent, S. Watanabe, A. A. Nugraha, J. Barker, and R. Marxer, "An analysis of environment, microphone and data simulation mismatches in robust speech recognition," *Computer Speech & Language,* 2016.
10. M. B. Mustafa, F. Rosdi, S. S. Salim, and M. U. Mughal, "Exploring the influence of general and specific factors on the recognition accuracy of an ASR system for dysarthric speaker," *Expert Systems with Applications,* vol. 42, pp. 3924-3932, 2015.
11. M. Cutajar, E. Gatt, I. Grech, O. Casha, and J. Micallef, "Comparative study of automatic speech recognition techniques," *IET Signal Processing,* vol. 7, pp. 25-46, 2013.
12. M. Steven Macaluso, M. Joseph Orange, and M. Robert Teasell, "Aphasia and Apraxia," 2016.
13. S. K. Bhogal, R. Teasell, and M. Speechley, "Intensity of aphasia therapy, impact on recovery," *Stroke,* vol. 34, pp. 987-993, 2003.
14. O. Saz, S.-C. Yin, E. Lleida, R. Rose, C. Vaquero, and W. R. Rodríguez, "Tools and technologies for computer-aided speech and language therapy," *Speech Communication,* vol. 51, pp. 948-967, 2009.
15. M. Kirmess and L. M. Maher, "Constraint induced language therapy in early aphasia rehabilitation," *Aphasiology,* vol. 24, pp. 725-736, 2010.
16. M. L. Albert, "Treatment of aphasia," *Archives of Neurology,* vol. 55, pp. 1417-1419, 1998.
17. N. Helm-Estabrooks, M. L. Albert, and M. Nicholas, *Manual of aphasia and aphasia therapy*: Pro-ed Austin, TX, 2004.
18. W. M. E. van de Sandt-Koenderman, "Aphasia rehabilitation and the role of computer technology: Can we keep up with modern times?," *International Journal of Speech-Language Pathology,* vol. 13, pp. 21-27, 2011.
19. L. Besacier, E. Barnard, A. Karpov, and T. Schultz, "Automatic speech recognition for under-resourced languages: A survey," *Speech Communication,* vol. 56, pp. 85-100, 2014.
20. J. Manikandan and B. Venkataramani, "Design of a real time automatic speech recognition system using modified one against all SVM classifier," *Microprocessors and Microsystems,* vol. 35, pp. 568-578, 2011.
21. A. Acero, *Acoustical and environmental robustness in automatic speech recognition* vol. 201: Springer Science & Business Media, 2012.
22. R. K. Aggarwal and M. Dave, "Acoustic modeling problem for automatic speech recognition system: conventional methods (Part I)," *International Journal of Speech Technology,* vol. 14, p. 297, 2011.
23. X. Huang, A. Acero, H.-W. Hon, and R. Foreword By-Reddy, *Spoken language processing: A guide to theory, algorithm, and system development*: Prentice hall PTR, 2001.
24. K. Gupta and D. Gupta, "An analysis on LPC, RASTA and MFCC techniques in Automatic Speech recognition system," in *Cloud System and Big Data Engineering (Confluence), 2016 6th International Conference*, 2016, pp. 493-497.
25. M. Anusuya and S. Katti, "Front end analysis of speech recognition: a review," *International Journal of Speech Technology,* vol. 14, pp. 99-145, 2011.
26. C. Kim and R. M. Stern, "Power-normalized cepstral coefficients (PNCC) for robust speech recognition," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP),* vol. 24, pp. 1315-1329, 2016.
27. Z. Tüske, P. Golik, R. Schlüter, and H. Ney, "Acoustic modeling with deep neural networks using raw time signal for LVCSR," in *Interspeech*, 2014, pp. 890-894.
28. M. H. Mohamad Jamil, S. Al-Haddad, and C. Kyun Ng, "A flexible speech recognition system for cerebral palsy disabled," *Informatics Engineering and Information Science,* pp. 42-55, 2011.

29. A. Garg and P. Sharma, "Survey on acoustic modeling and feature extraction for speech recognition," in *Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on*, 2016, pp. 2291-2295.
30. G. A. Saon and H. Soltau, "Method and system for joint training of hybrid neural networks for acoustic modeling in automatic speech recognition," ed: Google Patents, 2017.
31. P. P. Marsal, S. Pol, A. Hagen, H. Bourlard, and C. Nadeu, "Comparison and combination of RASTA-PLP and FF features in a hybrid HMM/MLP speech recognition system," in *INTERSPEECH*, 2002.
32. S. M. Mirhassani and H.-N. Ting, "Fuzzy-based discriminative feature representation for children's speech recognition," *Digital Signal Processing,* vol. 31, pp. 102-114, 2014.
33. S. Hegde, K. Achary, and S. Shetty, "Feature selection using Fisher's ratio technique for automatic speech recognition," *arXiv preprint arXiv:1505.03239,* 2015.
34. G. Garau and S. Renals, "Combining spectral representations for large-vocabulary continuous speech recognition," *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 16, pp. 508-518, 2008.
35. C. Ittichaichareon, S. Suksri, and T. Yingthawornsuk, "Speech recognition using MFCC," in *International Conference on Computer Graphics, Simulation and Modeling (ICGSM'2012) July*, 2012, pp. 28-29.
36. S. Xue, O. Abdel-Hamid, H. Jiang, L. Dai, and Q. Liu, "Fast adaptation of deep neural network based on discriminant codes for speech recognition," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP),* vol. 22, pp. 1713-1725, 2014.
37. S. Romdhani, "Implementation of DNN-HMM Acoustic Models for Phoneme Recognition," 2015.
38. P. Pujol, S. Pol, C. Nadeu, A. Hagen, and H. Bourlard, "Comparison and combination of features in a hybrid HMM/MLP and a HMM/GMM speech recognition system," *IEEE Transactions on Speech and Audio processing,* vol. 13, pp. 14-22, 2005.
39. E. Zarrouk, Y. B. Ayed, and F. Gargouri, "Hybrid continuous speech recognition systems by HMM, MLP and SVM: a comparative study," *International Journal of Speech Technology,* vol. 17, pp. 223-233, 2014.
40. M. L. Seltzer, D. Yu, and Y. Wang, "An investigation of deep neural networks for noise robust speech recognition," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013, pp. 7398-7402.
41. O. Abdel-Hamid, L. Deng, and D. Yu, "Exploring convolutional neural network structures and optimization techniques for speech recognition," in *Interspeech*, 2013, pp. 3366-3370.
42. Y. Zheng, "Acoustic modeling and feature selection for speech recognition," Citeseer, 2005.
43. S. O. C. Morales and S. J. Cox, "Modelling errors in automatic speech recognition for dysarthric speakers," *EURASIP Journal on Advances in Signal Processing,* vol. 2009, p. 308340, 2009.
44. F. Rudzicz, "Phonological features in discriminative classification of dysarthric speech," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, 2009, pp. 4605-4608.
45. S. R. Shahamiri and S. S. B. Salim, "Artificial neural networks as speech recognisers for dysarthric speech: Identifying the best-performing set of MFCC parameters and studying a speaker-independent approach," *Advanced Engineering Informatics,* vol. 28, pp. 102-110, 2014.
46. P. D. Green, J. Carmichael, A. Hatzis, P. Enderby, M. S. Hawley, and M. Parker, "Automatic speech recognition with sparse training data for dysarthric speakers," in *INTERSPEECH*, 2003.
47. M. Hariharan, C. Y. Fook, R. Sindhu, A. H. Adom, and S. Yaacob, "Objective evaluation of speech dysfluencies using wavelet packet transform with sample entropy," *Digital Signal Processing,* vol. 23, pp. 952-959, 2013.
48. S. A. Alim, N. K. A. Rashid, W. Sediono, N. Wahidah, and N. Nur, "LPC and its derivatives for stuttered speech recognition," *Jurnal Teknologi (Sciences & Engineering) 77: 18 (2015) 11–16,* vol. 77, pp. 11-16, 2015.
49. P. A. Heeman, R. Lunsford, A. McMillin, and J. S. Yaruss, "Using Clinician Annotations to Improve Automatic Speech Recognition of Stuttered Speech," *Interspeech 2016,* pp. 2651-2655, 2016.
50. O. C. Ai, M. Hariharan, S. Yaacob, and L. S. Chee, "Classification of speech dysfluencies with MFCC and LPCC features," *Expert Systems with Applications,* vol. 39, pp. 2157-2165, 2012.
51. D. Le and E. M. Provost, "Improving Automatic Recognition of Aphasic Speech with AphasiaBank," *Interspeech 2016,* pp. 2681-2685, 2016.