ISSN 2278-3091
Volume 9, No.5, September - October 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse123952020.pdf
https://doi.org/10.30534/ijatcse/2020/123952020

Remote-Controlled Car Based on Speech Recognition

H. N. M. Shah[1], J. T. Hoo[2], Z. Kamis[3], A. Ahmad[4], M. F. Abdollah[5]
[1,2,3,5]Centre for Robotics and Industrial Automation (CeRIA), Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka (UTeM), Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia.
[4]Faculty of Electrical and Electronic Engineering Technology, Universiti Teknikal Malaysia Melaka (UTeM), Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
hnizam@utem.edu.my

ABSTRACT

The concept of this project is to design speech recognition system that able to recognize the five basic command such as forward, reverse, left, right and stop. The five basic command is the command that normally used in car. The project more focused on the software part which is the speech recognition system. This speech recognition system was not recognizing a lot of words but only recognized the isolated words. The speech recognition is implemented in a robotic car to demonstrate the output result of the system. The microphone from the computer is used to give the input to the system. Convolutional Neural Network is used as feature extraction and feature classification of project. The software of the system contains training phase and testing phase. This speech recognized can works in every person, but the accuracy of speech recognition is different based on the pronunciation of the commands by the person. The overall accuracy of this speech recognition system recorded with 91.34% in a quiet room, 90.65% in a room during raining day and 77.98% on the roadside.

Key words : speech recognition, feature extraction, feature classification, cnn, speaker independent, isolated speech

1. INTRODUCTION

In this era of globalization, vehicles become one of the most important things for our daily life. Mohammed showed that 1.24 million death every year is causes by road traffic accidents especially 85% of death is occur in country under developing [1]. Many accidents are occurred when the traffic is jam. Traffic jam are physical draining, strain the driver nerves, and make driver more impatient. Their judgement when driving while be affected. A research from Loh stated that the speech recognition can be implement to the car [2]. A system that can help the driver for driving is design and lot of data collection and analysis needed to be done to succeed the recognition system.

Every human voice is entirely unique not only because of the actual shape and size of an individual's vocal cord but also due to the size and shape of the vocal tract. The human voice is one of the factors that will affect the accuracy of the speech recognition system. This will result in the accuracy of the speech recognition system being decreased. The difference of gender, nationality and race will influence the accuracy of the system due to different personal timbre. Kaur stated that every person have their own style of speaking [3]. It is found that a lot of research about speaker dependent speech recognition is the hottest trend in this recent. This may be even more complex configuration and analyses need to be done to succeed the recognition system.

Environment sound situation is giving impact to the accuracy of the speech recognition system. Le Prell stated that the major issue in communication for human and computer in a noisy environment is the background noise [4]. The different situation or environment have different type of background noise. The accuracy of the speech recognition system will be affected by the background noise. To collect accurate result, a lot of data collection and analysis need to be done under different situation.

After that, the time required to train the speech recognition and its accuracy plays an important role in determine the performance of the speech recognition system. That contains a lot of research about the effect of convolution layer to the performance of the speech recognition system. Saon stated that convolution layer may improve the speech recognition accuracy [5]. A research from Shea stated than the convolution layer will giving impact to the performance of the speech recognition system in terms of time required to trained and its recognition accuracy [6]. Therefore, a lot of data collection and analysis needed to done to improve the performance of the speech recognition system.

The main objective of this paper is to design robotic car using speech recognition system, investigate the relationship of convolutional layer and the performance of the trained network and evaluate the accuracy of the speech recognition system in different situation.

2. LITERATURE REVIEW

2.1 Type of speech

In speech recognition system, there are subdivided into three type of speech signals which are isolated speech, connected speech and continuous speech. Isolated speech

needed isolated word at a time. This type of speech needs the user to pause between pronouncements and isolated speech recognition system will have "Listen" or "Not-Listen" states. The noise on both side of the window contains very less noise [7]. Connected speech is a type of speech that between every word pronounced by human has a minimum pause. The connected word-observation sequence is contrasted with all possible regions of stored word-pattern in the word-level acoustic modelling [8]. Continuous speech is a kind of speech that has no gap between the words. Continuous speech is the most speech that normally being said by the human. There is a unique challenge for speech recognition to estimate the start and end of the pronounce by continuous speech. It is also difficult to create an algorithm to recognize the content of the speech.

Speech recognition contains two main categories which are speaker dependent and speaker independent. Speaker dependent is depend on the specific speaker. The recognition result is more accurate to the speech from the users that train the system while other result's speech will output less accurate result. This system is easier to implement, and the cost is low [9]. Beside that, speaker independent system is designed to recognize everyone voice so no training process is required. Different kinds of speech patterns, target word inflection and enunciation needed to respond by this system [10]. Although this type of speech recognition system can use for many speakers, but this system is difficult to implement, and price is higher. The speaker independent speech recognition system deals with fact by limiting the grammars while speaker dependent is likely to correctly to recognize what a speaker said due to smaller list of recognized words.

2.2 Feature extraction

Feature extraction is used to decrease the amount of data and selected the important feature from the selected features.[11-14]

There is some feature extraction technique had been widely used, for example Mel Frequency Cepstral Coefficient (MFCC), Linear Predictive Coding (LPC), Linear Predictive Cepstral Coefficient (LPCC), Perceptual Linear Prediction (PLP), Relative Spectral Perceptual Linear Prediction (RASTA-PLP) and Convolutional Neural Network (CNN).

MFCC considers human perception for sensitivity at appropriate frequencies by converting the conventional frequency to Mel Scale. The filter bank is applied by a Discrete Cosine Transform (DCT) to maintain result coefficients while discarding the rest because rapid changes in the filter bank coefficient and are useless in understanding the expression of their knowledge.

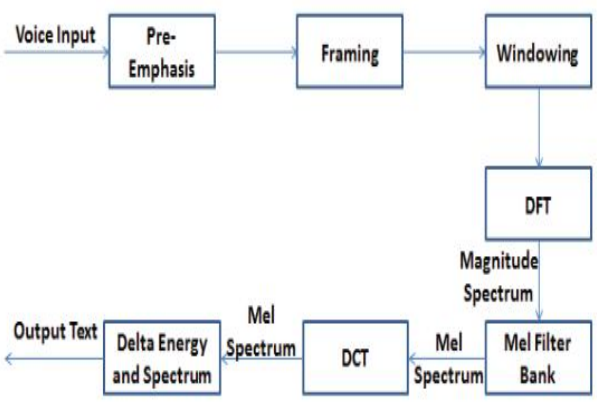

Figure 1: Block Diagram of MFCC [11]

Linear Predictive Coding (LPC) is a model based on approximation of the spectrum, which simplified the prediction of the vocal tract response from speech signals [15,16]. LPC works well in quiet, but not so in noisy environment. Pitch Period, Speech Frame Energy, Formant, bandwidth, and Short Time Spectrum are the parameters for speech signals.

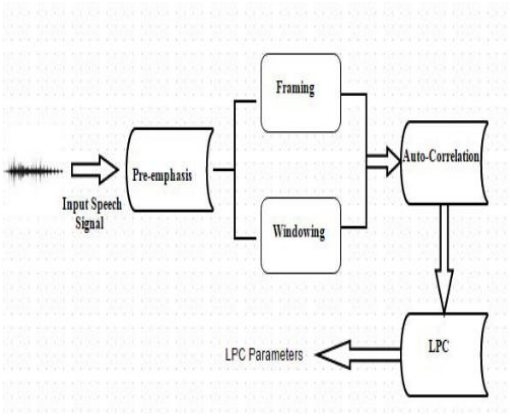

Figure 2: Block Diagram of LPC [13]

Perceptual Linear Prediction (PLP) method is similar with to LPC but PLP describes the psychophysics of human hearing more clearly by using logarithmic amplitude compression with critical-band power spectrum [17].

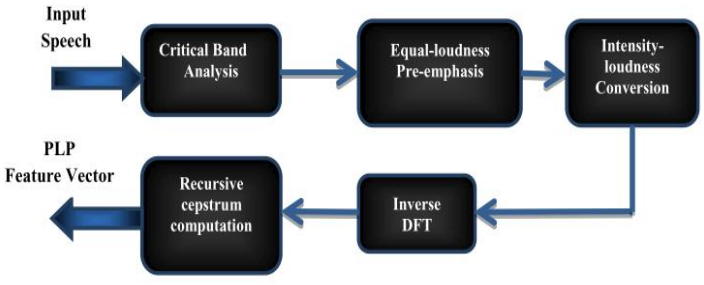

Figure 3: Block Diagram of PLP

2.3 Feature Classification

In acoustic phonetic approach, speech sounds where been searching and providing appropriate labels to recognize the speech. A set of acoustics properties in the speech characterized the phonemes which is finite and distinctive.

Hidden Markol Model (HMM) is popular when dealing with a wide range of time series data. This technique is a double-layered finite state process, with hidden Markovian process that is not observable to control the classification and selection of recognition process [18,19].

Dynamics Time Warping (DTW) technique is an algorithm that calculated the shortest distance between spoken word and reference templates [20]. DTW can analyses any data that can turned to linear representation. The important application is the automatic speech recognition (ASR) where DTW used to deal with different speaking speed. DTW algorithm implemented to calculate the shortest distance between feature of template and word input. The word detected based on the least value of calculated scores. Utterance of a same word will have different speech pattern because impossible to spoke the words with same rate each time.

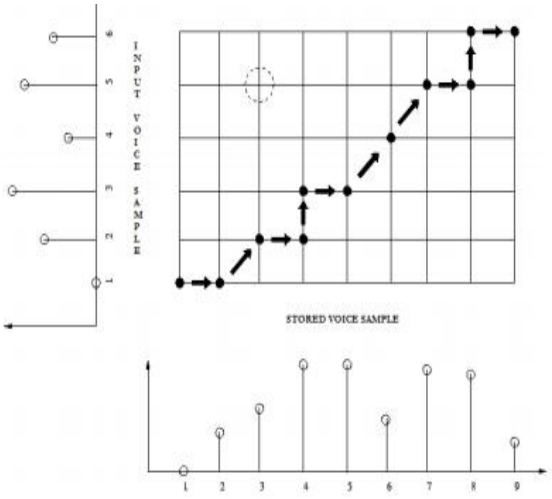

Figure 4: Example DTW of two voice samples [18]

Vector Quantization (VQ) technique is a traditional signal processing quantization technique that enables the stimulation of probabilities density functions by prototype vector distribution. VQ works by dividing a huge set of vectors into groups. VQ codeword is a centroid point that used to represent each vector group. Codebook is a set of codewords group. It contains many advantages, such as speed up the recognition process, can using for lightweight practical use and ease of implementation. But due to its advantages, VQ can lead to data loss when data compression. VQ can be used for speech recognition and speaker recognition system [21].

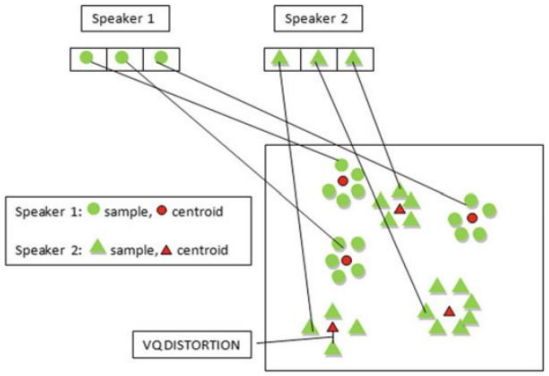

Figure 5: Example of VQ

CNN contains many attractive advancements, such as weight sharing, convolutional filters, and pooling. Therefore, CNN can achieve a great performance in Automatic Speech Recognition. Figure 6 showed that CNN consists of multiple layers of hidden layer which contains convolutional layers, pooling layers, and activation function. Each layer contains their own function such as convolutional layers used to filter the feature while pooling layer used to lower the size of the output data. There contains of enrolled filters in convolutional layers. The enrolled filters separate the image to specific size. Based on the research from Abdel-Hamid, the research showed that CNN have shown effectiveness in speech recognition [22]. Network training is a weight adjustment in neural network to create a network for predicting gradient descent method. Vishal showed that advanced architecture and optimized training can be used to train the CNN so it can achieve approximate human hearing level [23]. Non-linear function have been used to process the low-level data directly. CNNs are capable of learning high-level features with high complexity and abstraction. The heart of a CNN is the pooling because it able to reduce the dimensionality of a feature map. Back-propagation algorithm is used to perform the joint training of feature stage and classification stage.

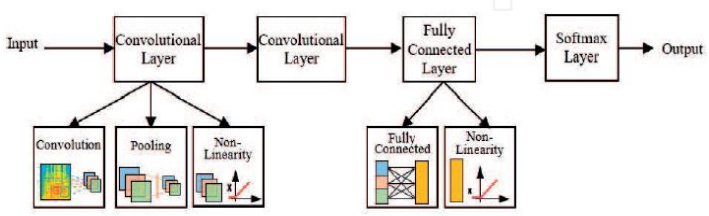

Figure 6: Basic structure of CNN

2.4 Type of Optimizer

Optimizer plays an important role for Convolutional Neural Network. The function of the optimizer is to update the weight parameters to minimize the loss function. There contains of many types of optimizer. The most common used

optimizer is Adaptive Gradient Algorithm (Adagrad), RMSProp and Adaptive Moment Estimation (Adam).

Adagrad is an adaptive learning rate algorithm. Adagrad choose the learning rate to the parameters. This type of optimizer works well with sparse gradient.

$$\theta_{t+1} = \theta_t - \alpha . g_t / \sqrt{\sum_i^t g_t^2} \quad (1)$$

RMSProp means Root Mean Square Propagation. RMSProp used a moving average of squared gradient to control the learning rates. The learning rate adjust automatically, and different learning rate is chosen for different parameter.

$$\theta_{t+1} = \theta_t - \frac{n}{\sqrt{(1-\gamma)g_{t-1}^2 + \gamma g_t + \varepsilon}} . g_t \quad (2)$$

Adam optimizer is the combination of Adagrad and RMSProp that works fine in online and nonstationary things. It decreases the diminishing learning rates of Adagrad. Adam is very efficient but has very little memory.

$$\theta_{t+1} = \theta_t - \frac{n m_t}{\sqrt{v_t + \varepsilon}} \quad (3)$$

Isolated speech used as the type of speech to recognize by this project because isolated speech can provide the highest accuracy due to the both side of the window contains less noise. The pause between the word easy to recognize by the speech recognition system. After that, among speaker dependent and speaker independent, speaker independent is selected as the type of speech recognition system. Speaker independent system required zero training to the users. Car is a vehicle that share among the family member, the speaker independent system can achieve the objective about car sharing with no training requirement for the speech recognition system. The speaker independent is more common to use because the car manufacturer does not need to train the speech recognition system for every car for each customer. The most important is the co-driver can give command in specific emergency time to reduce the accident rate.

After that, Convolutional Neural Network is used as the feature extraction and feature classification method. CNN is the hottest topic that used for speech recognition in this recent although deep learning is commonly used in image processing [24]. Convolutional Neural Network able to supply us a primary method to communicate with computer. Both Microsoft and Google have developed their product by using Convolutional Neural Network algorithms. Gao stated that CNN can achieve speaker independent speech recognition system with high accuracy [25]. Convolutional Neural Network able to extract feature from a huge of audio data given and classify the feature without the supervise from human. The pooling layers from CNN reduce the spatial size to decrease the parameters so a huge amount of data can be used to train the network.

3. METHODOLOGY

Convolutional Neural Network is used as the method to extract and classify the feature from the speech. Fig 7 shows the structure of the speech recognition in this speech recognition system.

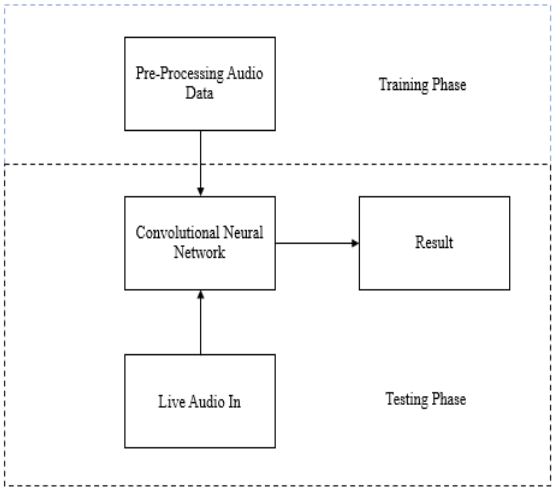

Figure 7: The structure of the speech recognition system

3.1 Pre-processing

Pre-processing the input data audio is the first step of this speech recognition system [26,27]. The audio dataset is distributed to training dataset, validation dataset and test dataset. Training dataset used to train the neural network while validation dataset and testing dataset used to test the performance of the trained network. Windowing, Framing, Fast Fourier Transform and Mel-Filter Bank is used to pre-processing [28] the audio file and transform it to audio spectrogram. Fig 8 is the example of spectrogram.

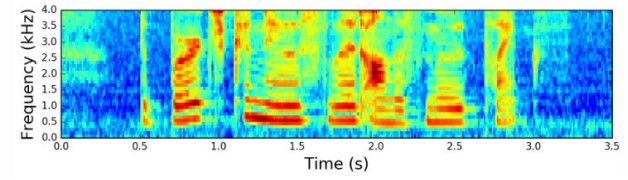

Figure 8: Spectrogram

3.2 Training Neural Network

Convolutional Neural is used as the method to recognize the speech command. Six 2-dimensional convolutional layers with size of [3 x 3] with different number of filters is created. The convolutional layer started with 32 filters and increase on

every layers. A batch normalization layer is placed between convolutional layer and Rectified Linear Unit (ReLU) layer to increase the training speed. A Max Pooling Layer with size [3 x 3] and stride [2 x 2] is placed after each convolutional layer to maximize the data of each region. Table 1 is the structure of convolutional Neural Network layer. The first layer is convolutional layer to extract the feature from an input speech while ReLU layer is applied to convert the negative value to zero. Max Pooling layers take the largest value from the rectified feature map and Fully Connected Layer is the layer for CNN to make prediction and classification on the recognition result. Figure 9 shows the flow of the CNN.

Table 1: Structure of the CNN layer

Layer(Type)	Size	Number of filter / strides
Convolution_1 + ReLU	3 x 3	32
Max Pooling_1	3 x 3	2 x 2 strides
Convolution_2 + ReLU	3 x 3	64
Max Pooling_2	3 x 3	2 x 2 strides
Convolution_3 + ReLU	3 x 3	128
Max Pooling_3	3 x 3	2 x 2 strides
Convolutional_4 +ReLU	3 x 3	128
Max Pooling_4	3 x 3	2 x 2 strides
Convolutional_5 + ReLU	3 x 3	128
Max Pooling_5	3 x 3	2 x 2 strides
Convolutional_6 + ReLU	3 x 3	128
Max Pooling_6	3 x 3	2 x 2 strides
Fully Connected Layer	-	-

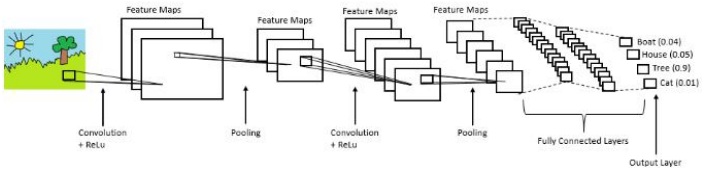

Figure 9: Flow of the CNN

Adam optimizer is used as the algorithm to train the neural network. Adam can optimize the performance of the trained network. The training process carry out with 25 epochs and each epoch means the entire training data pass through the network one time. The function of Adam optimizer is to update the network weights based on the training data. The training dataset is used in this section to train the neural network. All the training data is going through the deep learning algorithm and generate a convolutional neural network. The validation dataset is used to test the accuracy of the trained network.

3.3 Testing Phase

In this stage, user required to give speech command to the trained network system. The trained deep neural network will classify the command given by the user and recognize the speech command. A live speech detection figure is created to detect the speech command. The live audio will compute to live speech spectrogram before making classification with the trained network system. The trained network will classify the current spectrogram and make a predicted probability to recognize the speech command. A threshold probability with value 0.7 is implemented so the predicted probabilities with less than 0.7 will not declared. The predicted probabilities result which higher than 0.7 will be showed. Figure 10 shows the live speech command detection figure to detect the speech commands and transfer it spectrograms.

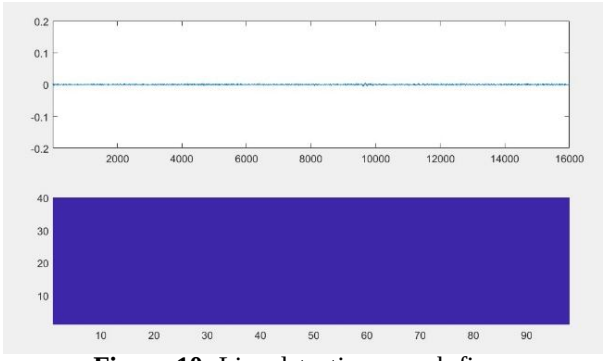

Figure 10: Live detection speech figure

4. RESULTS

4.1 Accuracy and training time for different number of convolutional layers

Convolutional Layer will give impacts to the validation accuracy and its required time to train the network. A experiment is carried out test the performance of the trained network by changing its number of convolution layer.

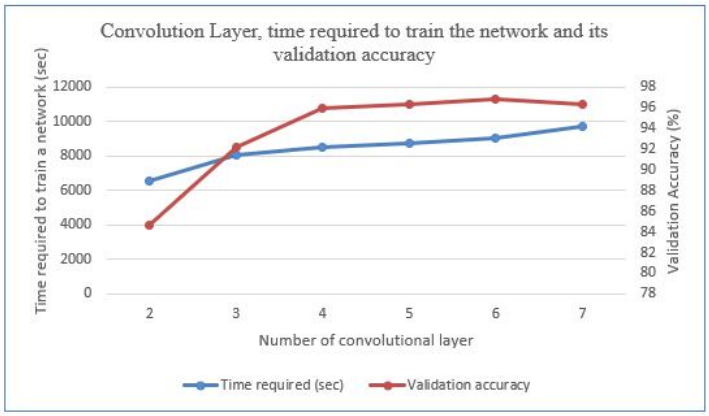

Figure 11: The number of convolutional layers, time required to train the network and the network validation accuracy for each layer

Based on the result obtained on Figure 12, the time required to train the network had increase as layers increased. This is because the train data set need to pass through more layer when the training process is carried on. The validation accuracy of the trained network is increasing from layer 2 to

layer 6 but slightly drop on layer 7. This is because overfitting of the data is occurred in this trained network.

4.2 Repeatability test for accuracy on multiple persons situation

Every person has their own unique sound. The accuracy of the speech recognition system is tested on multiple persons is proved that the system can function under different person. 30 trials of each speech command to obtain the accuracy.

$$\text{Accuracy} = (\text{correct} / \text{total}) \times 100\% \quad (4)$$

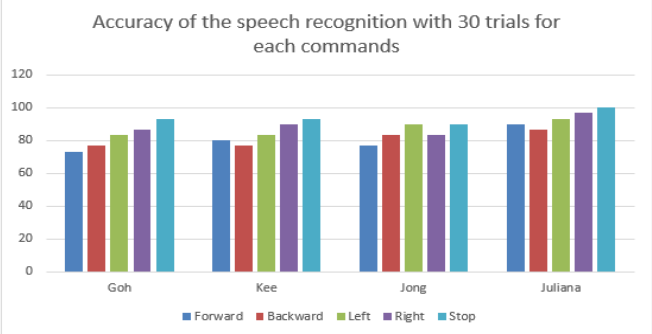

Figure 12: The comparison of the accuracy of each command in different person

Figure 12 proved the correct pronunciations of every command are important to obtain a correct recognition result. Four persons show different of accuracy. The 'forward' command gets lowest accuracy while 'backward' also get a low mark with 80.85% compare than 'left' command with 87.48%, 'right' command with 89.18% and 'stop' command with 94.15%. This is because the utterance of those 'forward' and 'backward' contains of two words which is 'for-ward' and 'back-ward'. The gap between the 'for' and 'ward' is too long, then the system will detect two word which is 'four' and 'ward' and the result shown is 'unknown'. The same condition also happens to 'backward' command.

4.3 Repeatability test for different environment situations

An experiment is conducted in three different environment situations to test performance of the speech recognition under different background noise. The location is set on a quiet room, a room during raining day and roadside.

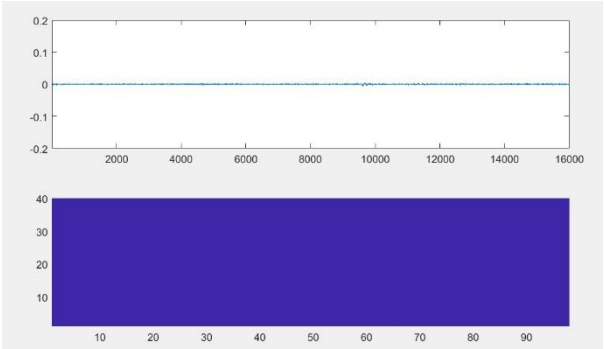

Figure 13: The background noise in quiet room

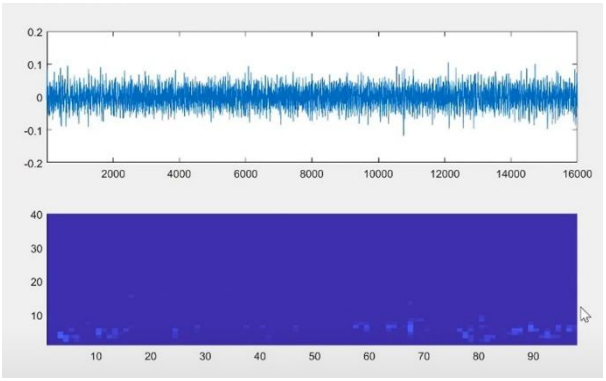

Figure 14: The background noise during raining day

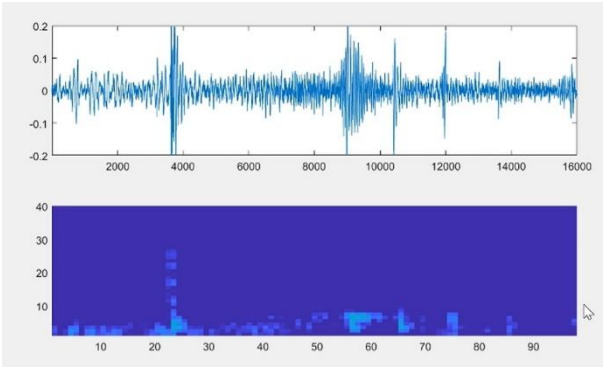

Figure 15: The background noise on the roadside

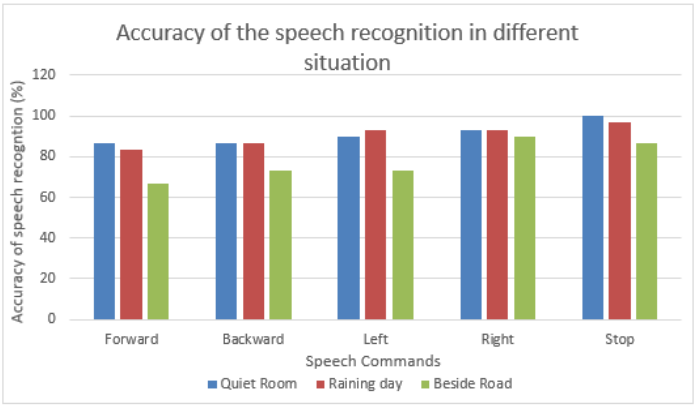

Figure 16: The comparison of each speech command in different environment

Figure 16 shows the recognition accuracy in a quiet room and raining day is almost the same. This is because the rainy sound has added to the background noise during the training phase of the network. The recognition accuracy at beside road is lower than the quiet room and raining day. The noise generated at beside the road is the highest among the three situations because road contains many noise generators such as car sound, car horn sound and wind sound. The heavy vehicle such as lorry pass through will produce a high volume of noise. The speech command given at that time is unable to

recognize by the system because the commands volume is cover by the lorry sounds. The speech recognition may declare an error output when the vehicle passes through with a sound like 'backward' command. That is a problem to insert the vehicle sound to the background noise during training phase because the vehicle sound different for every type of car.

5. CONCLUSION

As conclusion, Convolutional Neural Network can be used to develop the speech recognition system. The experiments done are matched with each of the objectives. The result showed that the more the number of convolutional layers, the higher the validation accuracy of the trained convolutional neural network until the trained network reached the overfitting limit. The findings proved that this speech recognition can recognize the command on multiple person situation, but the accuracy of the speech recognition depends on the correct pronunciation of each commands. The overall accuracy of this speech recognition system recorded with 91.34% in a quiet room, 90.65% in a room during raining day and 77.98% on the roadside. The accuracy of the speech recognition drops on the roadside due to the loud background noise from the vehicle pass through.

ACKNOWLEDGEMENT

The authors are grateful for the support granted by by Center for Robotics and Industrial Automation, Universiti Teknikal Malaysia Melaka (UTeM) in conducting this research through grant JURNAL/2018/FKE/Q00007 and Ministry of Higher Education.

REFERENCES

1. Mohammed, A. A., Ambak, K., Mosa, A. M. and Syamsunur, D. **A Review of the Traffic Accidents and Related Practices Worldwide**, Open Transp. J., vol. 13, no. 1, pp. 65–83, 2019.
2. Loh, C., Boey, K. L. and Hong, K. S. **Speech recognition interactive system for vehicle**, Proc. - 2017 IEEE 13th Int. Colloq. Signal Process. its Appl. CSPA 2017, pp. 85–88, 2017.
3. Kaur, J., Singh, A. and Kadyan, V. **Automatic Speech Recognition System for Tonal Languages: State-of-the-Art Survey**, Arch. Comput. Methods Eng., no. 0123456789, 2020.
4. Le Prell, C. G. and Clavier, O. H. **Effects of noise on speech recognition: Challenges for communication by service members**, Hear. Res., vol. 349, pp. 76–89, 2017.
5. Saon, G. and Picheny, M. **Recent advances in conversational speech recognition using convolutional and recurrent neural networks**, IBM J. Res. Dev., vol. 61, no. 4, pp. 1–10, 2017.
6. O'Shea, K. and Nash R. **An Introduction to Convolutional Neural Networks**, 2015.
7. Kandagal A. P. and Udayashankara, V. **Speaker Independent Speech Recognition Using Maximum Likelihood Approach for Isolated Words**, Int. J. Comput. Appl., vol. 7, no. 6, 2017.
8. Babu, T.A. and Kumar, P.R. **Prediction of Term Labor Using Wavelet Analysis of Uterine Magnetomyography Signals**, In Proceedings of International Conference on Computational Intelligence and Data Engineering, pp. 29-37, 2019.
9. Naziya S. and Deshmukh, R. R. **Speech Recognition System – A Review,** IOSR J. Comput. Eng., vol. 18, no. 04, pp. 01–09, 2016.
10. Kaberpanthi, N. and Datar, A. **Speaker Independent Speech Recognition using MFCC with Cubic-Log Compression and VQ Analysis**, Int. J. Comput. Appl., vol. 95, no. 26, pp. 33–37, 2014.
11. Sulaiman, M., Shah, H.N.M., Harun, M.H., Kazim, M. and Fakhzan, M.N. **Defect Inspection System For Shape-Based Matching Using Two Cameras**, Journal of Theoretical & Applied Information Technology, 61(2), 2014.
12. Marizan, S., Nizam, H., Shah, M., Mohamad Haniff, H., Lim, W.T., Kazim, , M.N.F.M. **A 3D gluing defect inspection system using shape-based matching application from two cameras**, 2013.
13. Shah, H.N.M., Sulaiman, M., Shukor, A.Z. and Ab Rashid, M.Z. **Recognition of butt welding joints using background subtraction seam path approach for welding robot**, International Journal of Mechanical & Mechatronics Engineering, 17(01), pp.57-62, 2017.
14. Shah, H.M., Sulaiman, M., Shukor, A.Z. and Ab Rashid, M.Z. **Vision based identification and detection of initial, mid and end Points of weld seams path in butt-welding joint using point detector methods**. Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 8(7), pp.57-61, 2016.
15. Wisesty, U. N. Adiwijaya, and Astuti, **W. Feature extraction analysis on Indonesian speech recognition system**, 2015 3rd Int. Conf. Inf. Commun. Technol. ICoICT 2015, pp. 54–58, 2015.
16. Liu, L., Li, W., Wu, X. and Zhou, B. X. **Infant cry language analysis and recognition: An experimental approach**, IEEE/CAA J. Autom. Sin., vol. 6, no. 3, pp. 778–788, 2019.
17. Gupta, H. and Gupta, D. **LPC and LPCC method of feature extraction in Speech Recognition System**, Proc. 2016 6th Int. Conf. - Cloud Syst. Big Data Eng. Conflu. 2016, pp. 498–502, 2016.
18. Younis, L. and Faisal, Y. **Speaker Dependent Speech Recognition in Computer Game Control**, Int. J. Comput. Appl., vol. 158, no. 4, pp. 32–38, 2017.
19. Gamit, M. R., Dhameliya, P. K. and Bhatt, N. S. **Classification Techniques for Speech Recognition : A Review**, Int. J. Emerg. Technol. Adv. Eng., vol. 5, no. 2, pp. 58–63, 2015.

20. Malode A. A. and Sahare, S. L. **An improved speaker recognition by using VQ & HMM**, IET Conf. Publ., vol. 2012, no. 624 CP, pp. 377–383, 2012.
21. Senthildevi, K. A. and Chandra, E. **Keyword spotting system for Tamil isolated words using Multidimensional MFCC and DTW algorithm**, Int. Conf. Commun. Signal Process. ICCSP 2015, pp. 550–554, 2015.
22. Mohan, B. J. and Ramesh Babu, N. **Speech recognition using MFCC and DTW**, Int. Conf. Adv. Electr. Eng. ICAEE, 2014.
23. Bansal, P., Imam, S. A. and Bharti, R. **Speaker recognition using MFCC, shifted MFCC with vector quantization and fuzzy**, Int. Conf. Soft Comput. Tech. Implementations, ICSCTI 2015, pp. 41–44, 2016.
24. Abdel-Hamid, O., Mohamed, A. R., Jiang, H., Deng, L., Penn, G. and Yu, D. **Convolutional neural networks for speech recognition**, IEEE Trans. Audio, Speech Lang. Process., vol. 22, no. 10, pp. 1533–1545, 2014.
25. Passricha, V. and Aggarwal, R. K. **Convolutional neural networks for raw speech recognition**. In From Natural to Artificial Intelligence-Algorithms and Applications. IntechOpen, 2018.
26. Dioses Jr, J.L. **AndroiDuino-Fan: A Speech Recognition Fan-Speed Control System utilizing Filipino Voice Commands**, International Journal of Advanced Trends in Computer Science and Engineering, 9(3), 2020.
27. Akour, M., Al Radaideh, K., Shadaideh, A. and Okour, O. **Mobile Voice Recognition Based for Smart Home Automation Control**, International Journal of Advanced Trends in Computer Science and Engineering, 9(3), 2020.
28. Shah, H.N.M., Ab Rashid, M.Z., Abdollah, M.F., Kamarudin, M.N., Lin, C.K. and Kamis, Z.. **Biometric voice recognition in security system**, Indian journal of Science and Technology, 7(2), p.104, 2014.