# SPEECH RECOGNITION INTERACTIVE SYSTEM FOR VEHICLES

Varsha Hotur[1], Usha N S[2]
[1]Asst.professor ECE Dept. MVJ College of engineering.
[2]Student of ECE Dept. MVJ College of engineering.

**Abstract**

**In this paper we have presented the design methodology of speech recognition interactive system for vehicles such as a car etc. speech recognition is the technology which enables a computer system to recognize the speech commands used or spoken by the user with the help of the device called microphone. Practically, increase the performance of speech recognition system like a number of commands or words used by the user, multiple users and noisy environment. nowadays, this speech recognition interactive system has been implemented to control the actions or features in cars such as controlling the sound system,lights by giving an user command and respective action will be performed in order to respond to user by comparing of current technology,there is a long history with the major innovative ideas for speech recognition system .this field has benefited from advances in big data concentration in learning and here using VW CAR_NET APP to provide an advantages like selfguiding purpose, management and to avoid more accidents in the city. CAR_ NET APP is directly connected with mobile to the car.if any misleading happens it will detect and gives a notification on the car screen or to the mobile device .notification that can be either text or voice message**

**Index terms: Speech recognition,vehicle system , pattern ,VW CAR_NET APP.**

## I. INTRODUCTION

This system implementation is about a voice command controlled system that is speech recognition interactive system for vehicles and this is used perform some actions or controls some actions such as open/close doors, switch on/off head lamp, turn on/off radio or multimedia system in car using voice commands by user that can be passenger and driver. The main purpose of implementing this system is to simplify the life of passenger and driver. This will be a user-friendly system .

The more concentration on driving will directly affect the danger of driver and passengers. There are a number of highly functional and complicated devices that need lots of concentration while driving vehicle or else accidents might occur. This system is for vehicles is implemented to reduce accidents while driving due to concentrating on other devices or other resources like during turn on/off of headlights etc. Also, implementing highly secured system and self-guiding, managing system using VW CAR-NET APP.car-net app will helps in when car is misleading or user searching for an address the system will guide the user to reach the destination by notifying the location information to mobile device. By using an user command the system without touching any control panels, any buttons it can perform the actions requested by user or driver.

few years ago launched automatic headlamp device all around the world by luxury automaker but in certain conditions of environment the headlamp sensor is not working properly .in the case of car passing through the underground tunnel but the headlamp will not work properly due to may be the environment is not dark enough.in order to solve this problem faced by driver , the speech recognition system is necessary to include the VW CAR-NET APP CONNECT in order to detect the location whether the location is underground or not.after giving notification driver is necessary to control the headlamp by driver itself in that situations rather than depending on sensors .

In this system using MATLAB as a development software and Mel frequency cepstral coefficients

(MFCC's) is used to extract the features of the signal provided by driver or passenger.

Furthermore, this speech recognition interactive system is expected to allow drivers without affecting their eye and hand movement during operation of car-mounted devices, by that way permitting them to carry out the tasks with ease. The complicated action such as dialling a cell phone is usually prohibited during driving.

## II. LITERATURE SERVAY

During the Speech Recognition process, the original speech will be distorted by the complex channel environment. Due to the channel distortion, the system may not able to capture the original signal produced by the user and therefore no output will be generated. [1] The research study with the topics of 'A new framework for robust speech recognition in complex channel' mentions that Channel distortion is the major factor that may degrade the performances of Automatic Speech Recognition (ASR) Systems [1].

A research study about 'An improved model of masking effects for robust speech recognition system' [2] explained that the performance of an automatic speech recognition system will drop dramatically in the presence of background noise unlike the human auditory system which is more adapted at noisy speech recognition[2]. In order to solve the problem, an auditory modeling algorithm is integrated into the feature extraction front-end for Hidden Markov Model (HMM) which is named later as LTFC to simulate the properties of human auditory system and be used in speech recognition system to enhance its robustness. Lastly, the proposed method sharpens the power spectrum of the signal in frequency and time domain.

In order to decrease the background noise produced by the external environment, optimizing the microphone array configuration may increase the quality of acquired speech as seen in a research study on a title of 'An evolutionary algorithm to optimize the microphone array configuration for speech acquisition in vehicles'. [3] Nowadays, there are variety of trending applications that use microphone array to perform speech acquisition in order to improve the quality of the acquired speech. The Multichannel speech enhancement based in spatial filtering is aimed. Actually the optimization of the filter coefficients has been the main focus in Beam Former design due to array configuration plays an important role in the quality of the speech acquisition system and it should be optimized.

In this mobile internet revolution world, browsing internet on mobile phone while driving is more and more commonly happened nowadays which may put the driver's safety at risk. Therefore, the in-car -speech interface to the web should be developed in order to provide driver with an intuitive and non-distractive driving environment as seen in a research study on 'Evaluation of speech-based HMI concepts for information exchange tasks: A driving simulator study. [4] Before the development of new speech dialog system (SDS), they examined the user's preferred interaction style and its influences on driving safety.

In this modern day, most of the information surrounding us are kept updating and there will be some new objects appeared each and every seconds. So the Self Learning skill must be available in speech recognition system so that it can be complete in the world nowadays. By using the interaction of speaker identification, speaker adaptation may achieve the unsupervised speaker tracking and automatic adaptation of human adaptation of the human-computer interface as seen in the research study by Tobias Herbig, Franz Gerl and Wolfgang Minker (2011).

The research study on the impact of accuracy and latency on mean opinion scores for speech recognition solutions had been done by James Scovell, Marco Beltman, Rina Doherty, Rania Elnaggar and Chaitanya Sreerama, 2015. The study had tested the effects of accuracy and latency on subjective ratings for a keyword triggered in speech solution. The researchers have developed the specialized software framework to trace the accuracy and latency of tasks and to understand the impact on user experience

In speech recognition interactive system[14].additionally using car-net app to provide maintenance and guidance also to provide more safety of driver.

In order to improve features in research study by Chee Yang Loh, Kai Lung Boey DQG Kai Sze Hong 2017 there is still lack of security.

In order to improve the security of driver in Speech recognition interactive system [14] adding **Volkaswagen CAR-NET APP-**

**CONNECT** to provide guidance and maintenance.

## III. DESIGN/ IMPLEMENTATION

During first stage of implementation, to code the algorithms to receive speech commands from user, using MATLAB as a development platform, to recognize the commands and to perform certain actions furthermore, a graphical user interface (GUI) is created to communicate with the user of vehicle system. In order to make sure that the working of speech recognition system of vehicle is correct, also database is needed to store all the speech data to match with the speech command spoken by the user of system.to perform matching process there are several algorithm and functions used in system. Algorithms used in this system are Mel frequency cepstral coefficients (MFCC's) and vector Quantization using linde-buzo-gray (VQLBG)

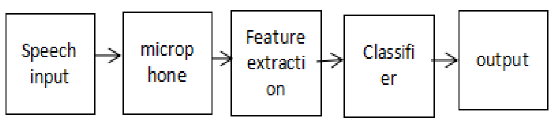


**Fig1..BLOCK DIAGRAM of speech recognition system.**

After the speech command is received by the system ,the signal frame is divided into frames before passing through the process of MFCC.there are currently 14 speech commands in the database due to hardware limitations and it will be further expanded when resources are sufficient to use.

Based on Fig. 1, there are several processes that are taken place in the processes of MFCCs. MFCC are used to extract the feature vector that contains all the information necessary about the linguistic message.

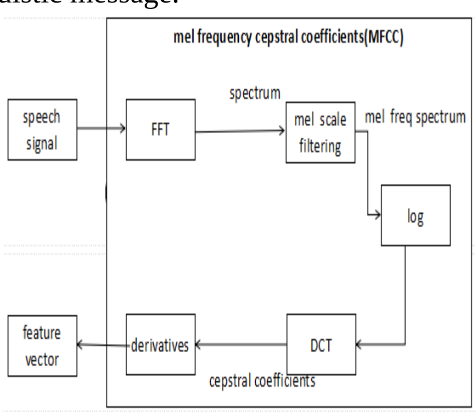


**Fig2. Block diagram of MFCC's process**

Moreover, MFCC mimics some parts of speech perception and speech production of human's ears and tries to eliminate speaker dependent characteristics by excluding the fundamental frequency and their harmonics. [7]

$$c^{(1)}_{\tau,k} = \left| \frac{1}{N} \sum_{j=0}^{N-1} f_j \, exp\left[ -i\, 2\pi \frac{jk}{N} \right] \right| \qquad k = 0, 1, ..., (N/2) - 1$$

Firstly, Discrete Fourier Transform is used to perform the computation of the frequency domain representation of the input signal [9].

$$c^{(2)}_{\tau,j} = \sum_{k=0}^{N/2-1} d_{j,k}\, c^{(1)}_{\tau,k} \qquad j = 0, 1, ..., N_d$$

Secondly, computation of the Mel-Frequency spectrum is used to mimic the human ears. The filter is inserted to mimic the human ear that act as human auditory system that uses the power over a frequency band as signal for further processing [10].

$$c^{(3)}_{\tau,j} = log(\, c^{(2)}_{\tau,j} \,) \qquad j = 0, 1, ..., N_d$$

Secondly, computation of the Mel-Frequency spectrum is used to mimic the human ears. The filter is inserted to mimic the human ear that act as human auditory system that uses the power over a frequency band as signal for further processing [10].

$$c^{(3)}_{\tau,j} = log(\, c^{(2)}_{\tau,j} \,) \qquad j = 0, 1, ..., N_d$$

Thirdly, this step computes the signal through logarithm in order to mimic the human perception of loudness due to experiments proved that humans perceive loudness in a logarithm scale [9].

$$F^{-1}\,\{log(\,F\,\{f_n\}\,)\}$$

Fourthly, the following step is used to compute the cepstral coefficient through eliminating the speaker dependent characteristics. Therefore, in order to take the lower order cepstral coefficients for further processing, the speaker dependent harmonics is suppressed [11].

$$c_\tau = \left[ c^{(4)}_{\tau,j} \;,\; \Delta c^{(4)}_{\tau,j} \;,\; \Delta\Delta c^{(4)}_{\tau,j} \right]$$

In the last process of the MFCC which is to compute the derivatives, the dynamic nature of speech is represented using first and second order derivatives of the cepstral coefficients in order to extend the feature vector [11].

After the speech is passed through the MFCC processes to extract the features of signal then pass through VQLBG [13].

Vector Quantization using Linde-Buzo-Gray model is the enhancement of the original Vector Quantization due to the latter is using a lossy data compression method based on the principle of block coding. Therefore, Vector Quantization using Linde-Buzo-Gray is used to perform data compression in order to reduce transmission bit rate while it is an easy and rapid algorithm that used to derive a good codebook by looking at the centroids of partitioned sets with the minimum distortion partitions. In additional, the first centroids are generated from all the training data by splitting procedure. By using the classification at each stage, the LBG able to find the nearest centroid to each vector and localize the optimization procedure

In order to improve the security of driver in Speech recognition interactive system[14] adding **Volkaswagen CAR-NET APP-CONNECT** to provide guidance and maintenance.

## IV. RESULTS

Table I shows the preliminary accuracy results of the Vehicle Speech Recognition Interactive System in real time environment with background noise. The accuracy obtained is around 78.57% as an average. By analyzing the results shown above, the overall speech commands that were inserted into the system to complete the matching process is persistent and no fatal error during the matching process.There are several testing carried out to test the system stability to handle speech command in different environment including recording in front of the outlet of air conditioner, the reason to carry out this experiment is because normally our vehicle interactive system will be located at the center console of the vehicle which near to the air conditioner vent that may create a lot of external noises when the air is blowing out from the air conditioner.

Moreover, others than the external noises that are generated in cars, there are also noises generated when the vehicle is idle such as vehicle engine noise, air conditioner compressor noise, road noise for those vehicles that do not have a good soundproofing. Therefore, the system is necessary to have the capability to handle such problems and several testing are carry out with this environmental condition to make sure that the system is capable to work in a practical environment.

After the action performed by the system for respective commands ,**VW CAR-NET APP** helps when misleading of car it will guide the car by giving notification to the user through mobile device or car screen .the notification either voice or text using app connected to car .and also helps in maintenance of car carefully.

Based on the results that were obtained from the worst condition environment, there are some weaknesses of the system for the speech commands of 'on' and 'off'. The system is difficult to extract the feature of these 2 words due to the features of these two words are quite similar to each other.

**TABLE I: THE OUTPUT RESULT OF THE SYSTEM**

| Speech Command Inserted | Speech Command Matched | Result Status (Right/Wrong) |
|---|---|---|
| Front Left Door ,Open | Front Left Door ,Open | Right |
| Front Left Door ,Close | Front Left Door ,Close | Right |
| Front Right Door ,Open | Front Right Door ,Open | Right |
| Front Right Door ,Close | Front Right Door ,Close | Right |
| Rear Left Door ,Open | Rear Left Door ,Open | Right |
| Rear Left Door ,Close | Rear Left Door ,Close | Right |
| Rear Right Door ,Open | Rear Right Door ,Open | Right |
| Rear Right Door ,Close | Rear Right Door ,Close | Right |
| **Headlamp, Switch On** | **Headlamp, Switch Off** | **Wrong** |
| Headlamp, Switch Off | Headlamp, Switch Off | Right |
| **Left Signal, Switch On** | **Left Signal, Switch Off** | **Wrong** |
| Left Signal, Switch Off | Left Signal, Switch Off | Right |
| Right Signal, Switch On | Right Signal, Switch On | Right |
| **Right Signal, Switch Off** | **Right Signal, Switch On** | **Wrong** |
| **Correct Speech Command Matched** | 11/14 | **78.57% (Efficiency)** |

Therefore, the user of the system needs to clearly speak the command so that the system can recognize them.if notification displayed on display system its easy to read without affecting to driving.

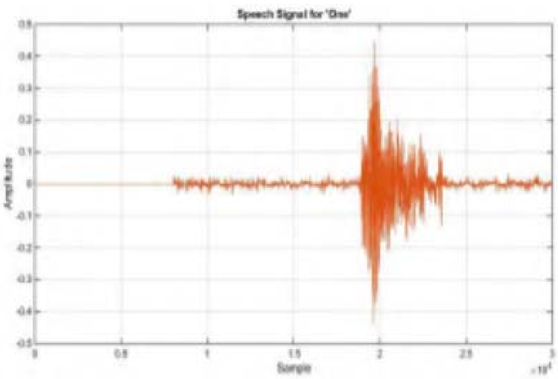


**Fig3. Speech signal for word "one"**

## V. CONCLUSION

In this speech recognition system for vehicles, the efficiency is 78.57%.and guidance and management is good .in most of the case the system cannot recognizes the speech command by user or drivers due to words of 'on ad 'off' which are quite similar to each other as studied in the result part .some improvements are done in order to improve the performance of the system such as multiple languages and larger database support will be added in the future to improve the performance of the system and also adding better featured app to maintain and guidance of vehicle .

## VI. REFERENCES

[1] Y. He, J. Han, T. Zheng and G. Sun, "A new framework for robust speech recognition in complex channel environments", *Digital Signal Processing*, vol. 32, pp. 109-123, 2014.

[2]P. Dai and I. Soon, "An improved model of masking effects for robust speech recognition system", *Speech Communication*, vol. 55, no. 3, pp. 387-396, 2013.

[3] D. Ayllón, R. Gil-Pita, M. Utrilla-Manso and M. Rosa-Zurera, "An evolutionary algorithm to optimize the microphone array configuration for speech acquisition in vehicles", *Engineering Applications of Artificial Intelligence*, vol. 34, pp. 37-44, 2014.

[4] H. Hofmann, V. Tobisch, U. Ehrlich and A. Berton, "Evaluation of speech-based HMI concepts for information exchange tasks: A driving simulator study", *Computer Speech & Language*, vol. 33, no. 1, pp. 109-135, 2015.

[5] T. Herbig, F. Gerl and W. Minker, "Self-learning speaker identification for enhanced speech recognition", *Computer Speech & Language*, vol. 26, no. 3, pp. 210-227, 2012.

[6] J. Scovell, M. Beltman, R. Doherty, R. Elnaggar and C. Sreerama, "Impact of Accuracy and Latency on Mean Opinion Scores for Speech Recognition Solutions", *Procedia Manufacturing*, vol. 3, pp. 4377-4383, 2015.

[7] K. IsmailaAdeniyi and O. Abdulhamid K., "Comparative Study on the Performance of Mel-Frequency Cepstral Coefficients and Linear Prediction Cepstral Coefficients under different Speaker's Conditions", *International Journal of Computer Applications*, vol. 90, no. 11, pp. 38-42, 2014.

[8] "The Speech Recognition Wiki", *Recognize-speech.com*, 2017. [Online]. Available: http://recognize-speech.com/feature-extraction/mfcc.

[9] H. Niemann, *Klassifikation von Mustern*, 1st ed. Springer-Verlag, 1983.

[10] K. Kido, *Digital Fourier Analysis: Advanced Techniques*, 1st ed., Springer, 2014.

[11] Chee Yang Loh, Kai Lung Boey DQG Kai Sze Hong

on Signal Processing & its Applications (CSPA 2017), 10 - 12 March 2017, Penang, Malaysia