

Driver Behaviour State Recognition based on Speech

Norhaslinda Kamaruddin^{*1}, Abdul Wahab Abdul Rahman², Khairul Ikhwan Mohamad Halim³, Muhammad Hafiq Iqmal Mohd Noh⁴

^{1,3,4}Advance Analytics Engineering Center (AAEC), Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40400 Shah Alam, Selangor, Malaysia

²Kulliyah of Information and Communication Technology, International Islamic University Malaysia, Gombak, Kuala Lumpur, Malaysia

*Corresponding author, e-mail: norhaslinda@tmsk.uitm.edu.my¹, abdulwahab@iium.edu.my²

Abstract

Researches have linked the cause of traffic accident to driver behavior and some studies provided practical preventive measures based on different input sources. Due to its simplicity to collect, speech can be used as one of the input. The emotion information gathered from speech can be used to measure driver behavior state based on the hypothesis that emotion influences driver behavior. However, the massive amount of driving speech data may hinder optimal performance of processing and analyzing the data due to the computational complexity and time constraint. This paper presents a silence removal approach using Short Term Energy (STE) and Zero Crossing Rate (ZCR) in the pre-processing phase to reduce the unnecessary processing. Mel Frequency Cepstral Coefficient (MFCC) feature extraction method coupled with Multi-Layer Perceptron (MLP) classifier are employed to get the driver behavior state recognition performance. Experimental results demonstrated that the proposed approach can obtain comparable performance with accuracy ranging between 58.7% and 76.6% to differentiate four driver behavior states, namely; talking through mobile phone, laughing, sleepy and normal driving. It is envisaged that such approach can be extended for a more comprehensive driver behavior identification system that may acts as an embedded warning system for sleepy driver.

Keywords: driver behavior state, speech emotion, silence removal, zero crossing rate, short term energy

Copyright © 2018 Universitas Ahmad Dahlan. All rights reserved.

1. Introduction

Department of Statistic Malaysia reported there are five principal causes of death in Malaysia; namely, ischaemic heart diseases (13.2%), pneumonia (12.5%), cerebrovascular diseases (6.9%), transport accidents (5.4%) and malignant neoplasm of trachea, bronchus and lung (2.2%) respectively in 2016 out of 85,637 deaths [1]. The statistics show an increase of 0.2% from 5.2% (4,196 deaths) in 2015 to 5.4% (4,640 deaths) in 2016 are yielded when data on transport accidents are compared. Transport accidents is also contributed to the third highest percentage of 7.4% for premature death in 2016. When comparing the data based on gender, male is more likely to have fatality with 3,923 deaths (7.5%) as opposed to female with 717 deaths (2.2%) in 2016. The trend has slight increase from 2015 with male recorded 3,496 deaths that is contrary to 700 deaths for female. The detail overview of principal cases of death in Malaysia for 2016 and 2015 is depicted in Figure 1. The trend of the involvement of male and female in a road traffic accident across European Union (EU) member states is consistent with Malaysia statistic. Male are more likely to report that they suffer from road traffic accident injury compared to female except for France where the proportion of men and women reporting a road traffic accident were the same [2]. It was also observed that the highest group to report that they have been injured in a road accident is the 35 – 44 years group. In a recent report produced by Ministry of Transport Malaysia, in 2016 alone, there are 521,466 accidents reported with injuries ranging from minor to fatal [3]. There is a 6.5% increase of number of accident reported (31860 accidents) as compared to 489,606 accidents in 2015. Figure 2 show the summary of casualties and deaths caused by road accidents in Malaysia from 2007 to 2016. Although it is obvious that the number of minor and severe injuries are declining, the statistic shows a consistent rise for death resulting from road accident. Hence, it is sound to find means and ways to study the cause of road traffic accident and such understanding can be extended to further reduce the number of accident.

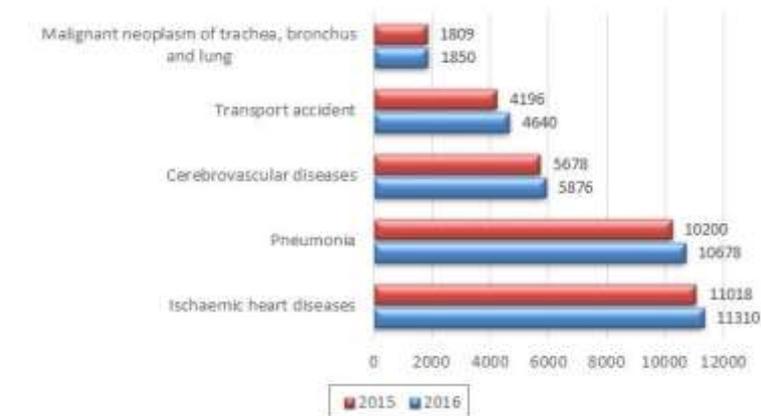


Figure 1. Statistics on Malaysia causes of death in 2015 and 2016 [1]

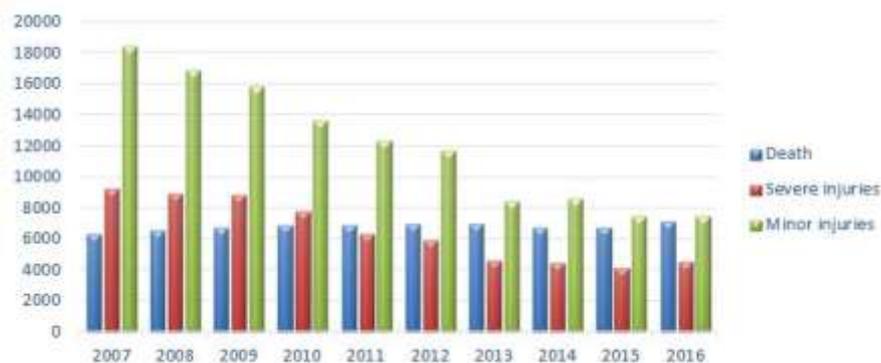


Figure 2. Statistics on Malaysia causes of death in 2015 and 2016 [3]

Traffic accidents typically are contributed by three main factors, namely; driver, vehicle and external environment [4]. In vehicle factor, lack of maintenance (i.e bald tires, bad brakes), mechanical failure (i.e vehicle age, overdue expiry date of spare parts) and design flaws (i.e manufacture malfunctions) are the most common reason why accident occurs. MIROS reported that vehicle defects contribute to 2% to the total cause of the accident [5]. The other 13% of the total cause of accident is recorded by road environment. Road environment consists situation such as hazardous road condition (i.e potholes, windy roads with no lines, steep shoulders, unsafe work zones for road repair, confusing road sign, defective traffic lights), road obstruction (i.e animal crossing, object on road) as well as weather and ambience (i.e fog, excessive rain, slick road, high wind, extreme differences in temperature, lighting as in sunrise and sunset) are the common environment aberrations that may cause accidents to happen. Subsequently, human factor is the most substantial reason that contributes 85% principal cause of road traffic crashes. According to Redhwan and Karim [6], fatigue, aggressive driving, sudden braking, following too close and exceeding speed limit are the common factors for driver behavior in Malaysia that leads to accident. Tawari [7] stated that there are different types of human factors such as distraction, drowsiness and emotion while driving. The potential distractions, are; drinking or eating, passenger disturbance, object in vehicle, using phone and other distractions [8]. According to Young and Regan [9], driver distraction happens when a driver is hindered from receiving information needed while doing the driving task. This is because of some event, activity, object or person that is within or outside the vehicle that switch the driver's attention from concentrating in the driving task. Kang [10] indicated that driver drowsiness and distraction have been the significant factors in many accidents because the driver awareness level and decision making capability are reduced which have negative effect to the driver itself. In

addition, human factor can also refer to human emotion influence which could be dangerous while driving. Psychologists observed that emotion influence directly by experience regarding how one feels about the object of surrounding [11]. Moreover, distraction, concentration, careless driving and loss of self-control while driving will also affect the emotion of driver [12-13]. Therefore, it is necessary to monitor the driver behavior and alert the driver when they are in distraction state to reduce the accident. Unsafe driving behaviors can be predicted in advance and this could lead to safe driving. According to Bayly [14], the amount of accidents would be reduced by 10% to 20% by monitoring and predicting the driver driving behavior states. For simplification, factors of road traffic accidents are illustrated in Figure 3.

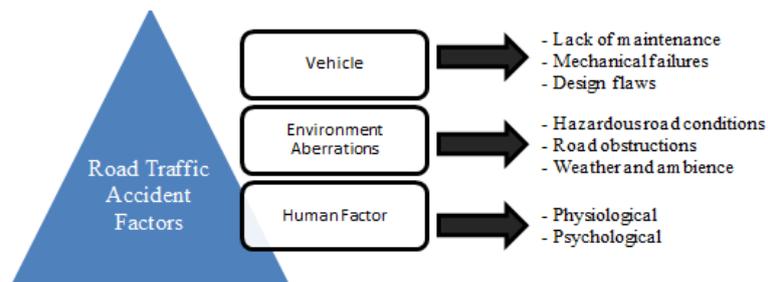


Figure 3. Road traffic accidents factors

Human factors can be simplified to four sub-sections to facilitate understanding; namely, physiological, psychological, behaviour and cognitive. Physiological factor refers to the aspects related to human well-being characteristics of normal functioning. Defects in this factor include fatigue, eye sight impairment / disorder, sleep deprivation, nutritional deficit, alcohol / drug / medicine intoxication or medical condition such as seizure, strokes, heart attack and false sensation of the sensory organs. The disability to judge based on previous experience, short attention span and low memory capacity are also falls into physiological factor. On the other hand, the psychological factors are concerned about the workings of the mind or psyche. It can be further segregated into motivation, perceptions, learning, beliefs and attitudes. For instance, acute stress, excessive emotion, lack of competence and skill, attitude (i.e negligence, arrogance, boldness, overconfidence), personality (i.e compromising, hardliner) and individual characteristics are common reason why accident may occur. In this paper, focus is given to psychological factor with emphasis on underlying emotion extracted from speech.

Speech can be used to measure driver behavior states (DBS) because speech carries underlying information that can differentiate between one DBS to another [12, 13]. However, speech data typically comprised of silence, voiced and non-voiced regions. Silence is observed when there is no speech is produced. It is different from unvoiced speech because the vocal cords are not vibrating resulting aperiodic and random speech waveform in nature. On the contrary, voiced speech produced a quasi-periodic speech waveform due to the air that flows from the lung that tensed of vocal chords. It is relatively high energy with less number of zero crossings present in the speech waveform [15]. Since for most practical cases, voiced region contains more information compared to the other two regions. Therefore, the silence and unvoiced are grouped together as silence region and need to be minimized. To complicate matters, background noise such as sound from vehicles engine, air conditioner, wind and others make the speech to noise ratio (SNR) relatively low resulting in difficulty to segregate between the data to be analyzed and artifacts. Hence, an automated tool to remove silence are developed using Energy and Zero Crossing Rate (ZCR). Such tool is useful in pre-processing a large number of data collected [16]. Although manual pre-processing data is the best, an automated tool can facilitate analysis in term of time, effort and monetary. In this paper, four different DBS are identified, namely; talking through mobile phone, laughing, normal and sleepy driving using Mel Frequency Cepstral Coefficient (MFCC) coupled with Multi Layer Perceptron (MLP) classifier. To explore the effect of silence removal, Short Time Energy (STE) and Zero Crossing Rate (ZCR) are used to truncate silence and unvoiced regions in order to make the data more compact. The aim of the paper is to enhance the driver behavior state (DBS)

recognition through the use of unvoiced and silence removal from the speech data signal thus improving the computational time and complexity.

This paper is organized in the following manner. Section 2 briefly described the Real-time Speech Driving (RtSD) dataset [12, 13] used for this work as well as feature extraction method and classifier employed. Section 3 demonstrates the use of Short Term Energy (STE) and Zero Crossing Rate (ZCR) to remove silence and unvoiced regions. The experimental set-up, results and discussion are provided in Section 4. In conclusion, Section 5 presented the summary and future work.

2. Real-time Speech Driving Dataset, Feature Extraction Method and Classifier

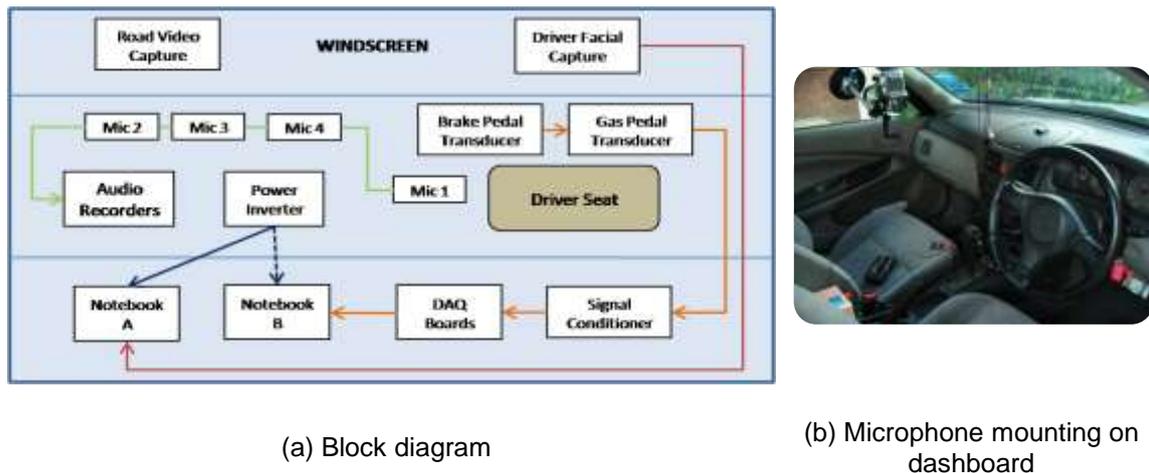
2.1. Real-time Speech Driving Dataset (RtSD Dataset)

Real-time Speech Driving Dataset (RtSD) [12-13] was collected by the Center for Computational Intelligence (C2iLAB) at the Nanyang Technological University Singapore. Eleven Singaporean and Malaysian drivers participated with age ranging from 20 to 54 years old with a minimum of five years driving experience. Each participant was required to drive covering 25.61 KM under differing traffic conditions and environments for approximately 60 minutes. Three microphones were placed all around the vehicles to record the ambient noise while one microphone was placed very close to the driver's mouth. Signals from all the four microphones were recorded and processed to achieve a cleaned speech. Each participant will have to go through three different driving conditions; a) normal driving condition listening to car radio with no interruption, b) heavy traffic with traffic lights and interrupted with interview being carried out in the vehicle, and c) no traffic light but heavy vehicles on road and the driver is required to make phone calls.

In this paper we investigate four different driving behaviour states (DBS); namely: talking through mobile phone, laughing, sleepy and normal driving condition. The talking through mobile phone distractor to the driver represents a medium stress DBS. Each driver was asked a simple questionnaire and needed to provide fast and accurate answers. The laughing DBS was captured when the driver was laughing while reading aloud the road directional signboards with the experimenter on board reading jokes. This is supposed to have a complementary effect to the stress induced in talking through mobile phone distractor. Normal condition driving will be used as the baseline since most of the time the driver will be driving under this condition. In addition, sleepy DBS was captured during the final phase of the driving exercise, when the driver is exhausted. The data was only analysed from selected drivers who complained of his/her sleepiness during the driving exercise especially towards the end of the driving task.

The data acquisition system comprises of recording a series of simultaneous data, such as: brake/gas pedal pressure signals, driver's facial expression and speech as well as road conditions, as shown by block diagram of Figure 4(a). Figure 4(b) shows the microphone mounted on the dashboard, the mounting of the video camera to record the road condition and the digital audio recorder recording noise in the vehicle. The aim for such comprehensive setting was to ensure a complete data collection for real-time driving data on an actual driving vehicle with various test subjects can be carried out. For the purpose of this study, only speech data is used for the analysis.

The driving route consists of six segments and the set of instructions for the driver to follow is divided into four phases. The route was planned to induce the driver to experience stress, distraction and frustration. All drivers were not familiar with the route thus periods of familiarization and rests were included in all analysis. For the study we had selected the route with an average amount of traffic at off-peak hours such that a typical daily commute and traffic can be simulated. This can also help the experimenters to control the situation better to ensure the safety of the driver. The detail description of the dataset can be found from [16].



(a) Block diagram

(b) Microphone mounting on dashboard

Figure 4. In-car data acquisition system

2.2. Mel Frequency Cepstral Coefficient (MFCC)

MFCC exploits the human auditory frequency response as in the cochlea which uses certain number of co-efficient filter bank and specific shape filter function. These features capture the perceptually most important parts of the spectral envelope of audio signals and translate the sound energy into the nerve impulses for the brain usage. Slaney MFCC implementation [17] of extracting 40 features from the speech signals was selected based on Ganchev et al. claim that Slaney's approach gives slightly better performance than others [18].

2.3. Multi-layer Perceptron (MLP) Classifier

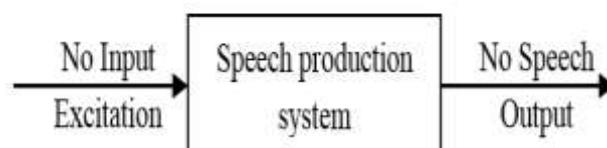
Multi Layer Perceptron is a feed-forward neural network trained with the standard back-propagation algorithm. MLP has the ability to find nonlinear boundaries separating the states. The complexity of MLP network can be changed by varying the number of hidden layers and the number of neuron unit in each layers. Given enough hidden unit and training data, MLP can approximate virtually any function to desired targets [19]. During training, error information is propagated back to the network to adjust the weight of each neuron and map the output with the most minimal mean square error. In this work, 1- and 2-layer MLP with 10, 20, 30 and 40 neurons are used to observe the DBS identification performance.

3. Silence Removal

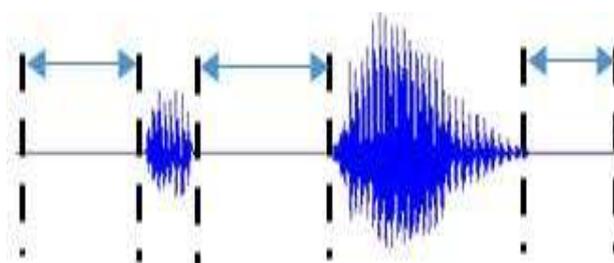
Real world data is often distorted due to the present of noise and artifacts that may cause linear or non-linear transformation of the original data. Analysing corrupted or distorted data may gives wrong results thus yielding wrong conclusion. Hence, it is imperative that data to be used for any analysis must be pre-processed to ensure it is free from noise and artifacts. A clean data is needed for the analysis to ensure the observation is derived from the correct data. Thus the raw data to be used by feature extraction and classification stages must first be preprocessed by removing all the noise and artifacts. In this work, we only focus on silence and unvoiced regions removal based since the noise produce by the vehicle is very obvious and the signal to noise ratio (S/N) is the lowest. In the vehicular environment the engine noise and other ambient noise will be maximum during the non-voice region, thus it will not be useful for our driving behavior analysis. For simplicity in this paper we use the silence and unvoiced regions interchangeably which refer to silence region.

In speech production, there are silence regions exists in between voiced and unvoiced speech signal. The silence region is characterized by the absence of any speech signal characteristics. Silence is essential for human to comprehend the speech but for our analysis it becomes redundant and need to be removed. Figure 5 depicts the silence region in the speech signal. During silence region signal occurrence, there is no excitation input and output to the vocal tract as shown in the block diagram of Speech with Silent Region of Figure 5(a). It has the lowest energy compared to unvoiced and voiced speech segments as shown in Figure 5(b). It is

also observed that there are relatively more number of zero crossings compared to unvoiced segment and no correlation among successive samples. In Figure 5(b), the arrows show the silence region between utterances.



(a) Block diagram



(b) Example of wave signal

Figure 5. Speech signal containing silence region

Two of the most common approaches to detect silence region are by employing the Zero Crossing Rate (ZCR) and Short Time Energy (STE). The ZCR can be defined as the signal rate changes of either positive or negative while the signal are transmitted [20]. It is a measure of number of times in a given time interval/frame that the amplitude of the speech signals passes through a value of zero. ZCR is very popular for Voice Activity Detection (VAD) to determine between voiced, unvoiced and silence region. Such ability is due to the fact that ZCR rate for unvoiced sounds and noise are usually higher than voiced sounds thus making it possible to detect the start and end point of unvoiced sounds.

The Short Term Energy (STE) can be defined as the energy with a short speech segment [21]. It is a simple and effective classifying parameter specially to differentiate between voiced and unvoiced sounds or silence because typically the voiced signal produces higher energy than silence. The silence region determined by STE will produced less energy than a certain threshold, which will then be truncated.

The combination of ZCR and STE (ZCR+STE) is used because in voiced speech, the STE values are much higher than in unvoiced speech and has higher zero crossing rate. The input signal is calculated based on ZCR and STE with different type of windowing and the output is processed by silence removal frame by frame using this output value. The calculation is based on number of frame and each frame is checked to determine whether silence region existed or not. The condition stated that if the maximum amplitude of original input is less than the maximum output, the signal will be truncated. In addition, if the minimum of energy is less than the threshold, it is considered as silence and the frame will be omitted. The ZCR+STE is calculated using window length of 200. The window length is the value which influence the detection of voiced, unvoiced and silence signal. Figure 6 shows the flow of the silence removal using ZCR+STE. In this work, Hamming window is used. The Z is the output of ZCR_STE function while S is the original speech. Framing is used to separate the Z and S signal frames of 0.01 seconds of frame. The number of frame is calculated by dividing the length of S with the frame length. Each of the 0.01 seconds of S and Z is then compared. If $\max(S) \geq \max(Z)$, the signal is appended and go to the next iteration. Otherwise, the frame is removed. Finally, the clean signal is returned (with removed unvoiced and silence).

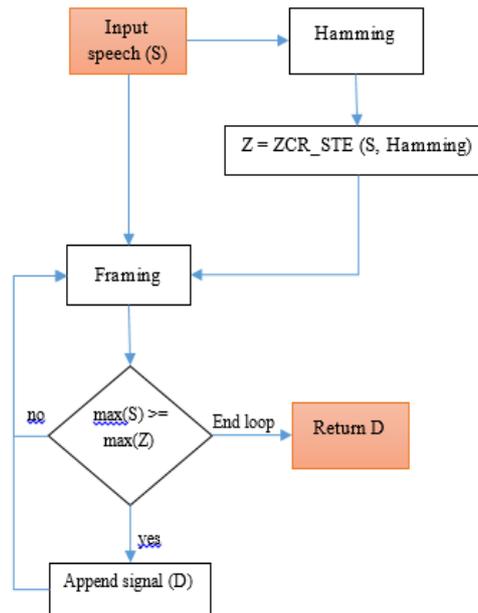


Figure 6. Silence region removal using ZCR+STE

4. Experimental Set-up, Result and Discussion

Once the data had been cleaned, MFCC feature extraction method and MLP classifier were employed for DBS identification. In this work, two types of data arrangements were used with different number of instances in the targeted DBS class, namely: a) talking-biased and b) even-distribution data arrangements. Talking-biased data arrangement comprised of 2000 instances of talking through mobile DBS with 666 instances for sleepy and laughing DBS respectively and the remaining 668 instances for normal driving DBS. This arrangement assumes in real life situation where more drivers will be talking through their headphones while driving simulating the medium level stress distractor. In addition this data arrangement also allow us to analyse aggravated drivers talking on the handphones and provided a free and unconstrained the way for the driver to responds. On the contrary, the even-distribution data arrangement consists of 1000 instances for the four studied DBS. It is hypothesized that the accuracy for talking through mobile DBS will be the most recognized DBS in the talking-biased data arrangement DBS recognition experiment whereas a more consistent performance among the four DBS should be observed in the even-distribution data arrangement experiment.

5-fold validation technique is employed using 80-20 rule. The data is segregated randomly in 5 folds where 80% of the data is used for training and the remaining 20% of the data is used for testing. The training-testing pairs are iteratively changed until the data are used completely. This is to ensure the classifier calculate the generalization of the data instead of memorization (using similar data for training and testing). 8 MLP networks architecture are implemented using 1 and 2 hidden layers with 10, 20, 30 and 40 neurons respectively. Figure 7 presents the identification performance using the talking-biased data arrangement using multiple MLP network architectures.

Results in Figure 7 illustrated that laughing DBS is consistently the least identified as compared to the other DBS with the lowest performance recorded using one hidden layer MLP with 10 neurons (11.41%) and the best performance using 2 hidden layers with 30 neurons (32.88%). The talking through mobile DBS results performed as expected with mean performance of 84.4% and 83.6% for one and two layers MLP respectively. It gives almost two times better than the accuracy of sleepy and normal DBS that yielded performance ranging between 35% and 49%. Hence, it shows that the size of instances in a class may affect performance and different MLP networks architecture may give different performance.

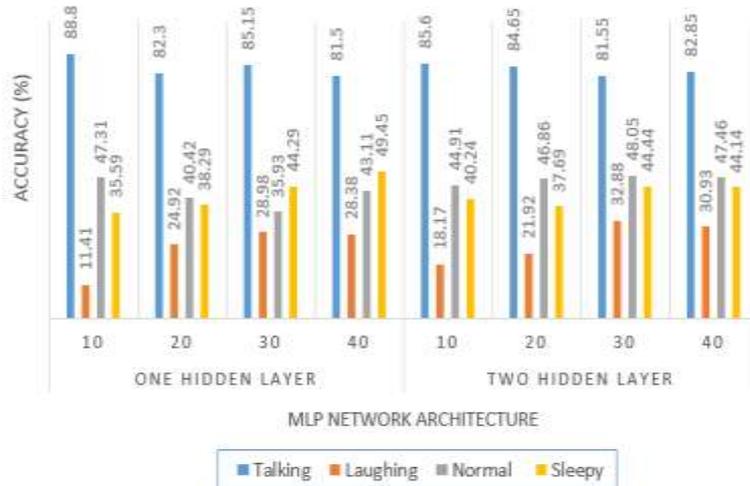


Figure 7. Silence region removal using ZCR+STE

Further analysis was conducted to determine the optimal MLP network architecture for even-distribution data arrangement DBS identification experiment. Figure 8 shows the overall mean performance of DBS identification result using talking-based data arrangement. It is observed that the highest accuracy is yielded when MLP 2 hidden layer with 30 neurons was used with 51.73% accuracy. Hence, such MLP network architecture will be employed for even-distribution data arrangement DBS identification experiment. The main reason of conducting this experiment is to note the changes in accuracy detected by having no bias in the number of instances in target classes

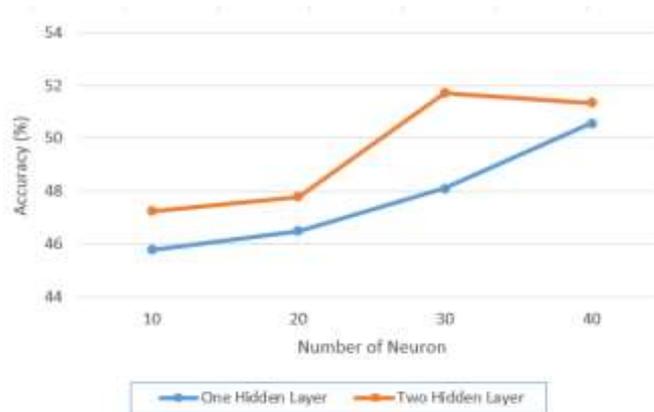


Figure 8. Overall mean performance of DBS identification result using talking-biased data arrangement

Table 1 illustrated identification results using the even-distribution arrangement data with highest DBS identified as normal driving (76.6%) followed with sleepy DBS (71.6%), talking through mobile DBS (59.2%) and lastly laughing DBS (58.7%). The overall mean performance recorded is 66.53% that is about 15% better than the best performance recorded using talking-biased data arrangement. The result is more distributed with variance of 13.4% indicating the potential of such approach to be implemented in recognizing different DBS.

Table 1. The DBS Identification Result Using Even-distribution Data Arrangement

Result	Target				Accuracy (%)
	Talking	Laughing	Normal	Sleepy	
Talking	592	208	37	77	59.2
Laughing	171	587	68	79	58.7
Normal	129	108	766	128	76.6
Sleepy	108	97	129	716	71.6

5. Summary and Conclusion

Recognising driving behavior state (DBS) can help reduce the road traffic accidents rate. Result from Figure 7 and 8 and Table 1 shows that it is possible to identify different DBS through speech and by removing the silence and non-voice region. Table I shows the potential of using the speech DBS system to recognize sleepy driver which can be very useful in identifying potential abnormal driving behavior that can cause accidents. It is also shown that due to the silence and unvoiced removal the speech data had been reduced to more than 50% of its original form thus improve the computational time required. Even for cases of talking on mobile phone can be recognize and differentiated with normal driving with 60% and 77% accuracy respectively.

This paper shows a preliminary work on DBS which can be enhanced further with more speech data and different DBS. Further works should be extended in term of feature extraction [22-23], optimal classifier architecture and more driving speech data. It is hope that such work can help reduced our traffic accidents making our road safer for both the drivers and pedestrians.

Acknowledgement

The authors would like to thank MARA University of Technology, Malaysia (UiTM) and Ministry of Education Malaysia (KPM) for providing financial support through the Fundamental Research Grant Scheme FRGS (600–RMI/FRGS 5/3 (106/2014)) to conduct the work published in this paper.

References

- [1] Department of Statistics Malaysia, Statistics on Causes of Death, Malaysia, 2017. Retrieved December 20, 2017 from <https://www.dosm.gov.my/v1/index.php?r=column/pdfPrev&id=Y3psYUI2VjU0ZzRhZU1kcVFMThGUT09>
- [2] Eurostat, Accidents and Injuries Statistics. Retrieved December 19, 2017 from http://ec.europa.eu/eurostat/statistics-explained/index.php/Accidents_and_injuries_statistics
- [3] Ministry of Transport, Transport Statistics Malaysia 2016, Retrieved December 19, 2017 from <http://www.mot.gov.my/my/Statistik%20Tahunan%20Pengangkutan/Statistik%20Pengangkutan%20Maysia%202016.pdf>
- [4] Department for Transport, Great Britain, Contributory Factors to Reported Road Accidents 2014, Retrieved on August 9, 2017 from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/463043/rrcgb2014-02.pdf
- [5] Malaysian Institute of Road Safety (MIROS), Road Facts, Retrieved August 9, 2017 from <https://www.miros.gov.my/1/page.php?id=17>
- [6] Redhwan AA, Karim AJ. Knowledge, Attitude and Practice towards Road Traffic Regulations among University Students, Malaysia. *The International Medical Journal of Malaysia*. 2010; 9(2): 29-44.
- [7] Tawari A, Trivedi MM. *Audio Visual Cues in Driver Affect Characterization: Issues and Challenges in Developing Robust Approaches*. The 2011 International Joint Conference on Neural Networks (IJCNN). San Jose. 2011: 2997-3002.
- [8] Stutts JC, Reinfurt DW, Staplin L, Rodgman EA. *The Role of Driver Distraction in Traffic Crashes*. Washington, DC: AAA Foundation for Traffic Safety, 2001.
- [9] Young K, Regan M. Driver Distraction: A Review of the Literature. *Editors I.J. Faulks, M. Regan, M. Stevenson, J. Brown, A. Porter & J.D. Irwin*, Distracted Driving. Sydney, NSW: Australasian College of Road Safety, 2007: 379-405
- [10] Kang HB *Various Approaches for Driver and Driving Behavior Monitoring: A Review*. International Conference on Computer Vision Workshops. Sydney, Australia. 2013: 616-623.

- [11] Clore G, JR Huntsinger, How Emotions Inform Judgment and Regulate Thought. *Trends in Cognitive Science*, 2007; 11: 393-399.
- [12] Kamaruddin N, Wahab A. *Driver Behaviour Analysis through Speech Emotion Understanding*. Intelligent Vehicles Symposium (IV), San Diego, California, USA, 2010:238-243.
- [13] Kamaruddin N, Wahab A. *Heterogeneous Driver Behavior State Recognition using Speech Signal*. The 10th WSEAS International Conference on System Science and Simulation in Engineering (ICOSSSE 2011), Penang, Malaysia, 2011:207-212.
- [14] Bayly M, Fildes B, Regan M, Young K. Review of Crash Effectiveness of Intelligent Transport Systems. *Emergency*. 2006; 3: 1-131.
- [15] Radmard M, Hadavi M, Nayebi M. A New Method of Voiced/Unvoiced Classification based on Clustering. *Journal of Signal and Information Processing*, 2011; 2(4):336-347.
- [16] Khalid M, Wahab A, Kamaruddin, N. *Real Time Driving Data Collection and Driver Verification Using CMAC-MFCC*. The International Conference on Artificial Intelligence (ICAL '08), Las Vegas, California, USA, 2008:219-224.
- [17] Slaney M. Auditory Toolbox. Ver 2. *Technical Report #1998-010*, Interval Research Corporation, 2007, Retrieved on June 13, 2017 from <http://cobweb.ecn.purdue.edu/~malcolm/interval/1998-010>
- [18] Ganchev T, Fakotakis N, Kokkinakis, G. *Comparative Evaluation of Various MFCC Implementations on the Speaker Verification Task*. The 10th International Conference on Speech and Computer (SPECOM 2005), 1, Patras, Greece, 2005: 191–194.
- [19] Zurada JM. *Artificial Neural System*. St Paul: MN:West Publishing. 1992.
- [20] Chen CH. *Signal Processing Handbook*, Electrical Computer Engineering, 51, New York: CRC Press, 1988.
- [21] Yang X, Tan B, Ding J, Zhang J, Gong J. *Comparative Study on Voice Activity Detection Algorithm*. International Conference on Electrical and Control Engineering (ICECE). Wuhan, China. 2010: 599-602.
- [22] Kamaruddin N, Wahab A. *CMAC for Speech Emotion Profiling*. The 10th Annual Conference of the International Speech Communication Association (INTERSPEECH 2009), Brighton, UK. 2009: 1991-1994.
- [23] Kamaruddin N, Wahab, A. *Human Behavior State Profile Mapping Based on Recalibrated Speech Affective Space Model*. The 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '12), San Diego, USA. 2012: 2021–2024.