Journal of Engineering Science and Technology
Vol. 12, No. 11 (2017) 3102 - 3119

A DESIGN FRAMEWORK FOR HUMAN EMOTION RECOGNITION USING ELECTROCARDIOGRAM AND SKIN CONDUCTANCE RESPONSE SIGNALS

KHAIRUN NISA' MINHAD*, SAWAL HAMID MD ALI,
MAMUN BIN IBNE REAZ

Department of Electrical, Electronic and Systems Engineering,
Faculty of Engineering and Built Environment,
Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia
*Corresponding Author: khairunnisa.minhad@siswa.ukm.edu.my

Abstract

Identification of human emotional state while driving a vehicle can help in understanding the human behaviour. Based on this identification, a response system can be developed in order to mitigate the impact that may be resulted from the behavioural changes. However, the adaptation of emotions to the environment at most scenarios is subjective to an individual's perspective. Many factors, mainly cultural and geography, gender, age, life style and history, level of education and professional status, can affect the detection of human emotional affective states. This work investigated sympathetic responses toward human emotions defined by using electrocardiography (ECG) and skin conductance response (SCR) signals recorded simultaneously. This work aimed to recognize ECG and SCR patterns of the investigated emotions measured using selected sensor. A pilot study was conducted to evaluate the proposed framework. Initial results demonstrated the importance of suitability of the stimuli used to evoke the emotions and high opportunity for the ECG and SCR signals to be used in the automotive real-time emotion recognition systems.

Keywords: Biomedical signal processing, Emotion recognition, Signal data acquisition protocol, Stimulus methods.

1. Introduction

Road accidents in Malaysia recorded RM78 billion lost with an average of RM1.2 million every year. In 2014, the estimated damage costs were RM9.3 billion, and the fatality rate was 24 for every 100,000 inhabitants.

Nomenclatures

bps	Signal bandwidth in unit bit per seconds
V	Device output range in unit Volt
z	Measurement of standard deviations from the population mean

Greek Symbols

α	Type I error of level of significance value
β	Probability of a Type II error
μ	Mean value (average) of a population
$\mu Siemen$	Amplitude of measured SCR signal in micro Siemen
σ^2	Variance value by square of each data and average the result

Human errors including risky driving, speeding recklessly, fatigue and driving under influence had contributed to 68% of road accidents [1]. Risky driving which associated with the human state of emotion including frustration, angry, and hostility are repeatedly linked to risky and aggressive driving [2, 3]. Emotion is defined as a spontaneous feeling or mental state of the individual that occurs without the individual's conscious effort but can be detected by the physiological changes of human body. Physiological signals originate from the activity of the autonomous nervous system (ANS) in the human brain and they cannot be triggered by any conscious or intentional control. Hence, suppressing or socially masking the emotion is impossible [4]. An integrated system built-in the vehicle is expected to provide an immediate and necessary response based on the identification of the driver's negative emotional state. This system can also support fatal accident investigations through the recorded emotional state presented during driving. Two physiological signals investigated in this work are electrocardiogram (ECG) and electrodermal activity (EDA) which focusing on the skin conductance response (SCR). Both ECG and SCR signals have close correlation to human emotions that can be obtained non-intrusively during the driving situations.

Electrocardiogram wave is used to capture and amplify tiny electrical activities detected on the skin and reported as one of the biological signals that produces the best classification of mental status [5]. Hiding emotions with regard to cardiac reactions is difficult and ECG reportedly exhibits emotion specific pattern after ANS ends in each of the four chambers of the heart [6]. In the regular activity of the heart, ECG reflects the health state of driver as well as checks the degree of operation of ANS through an analysis of heart rate variability (HRV). The electrodermal activity parameters that include both SCR and SCL are reported to indicate audience arousal during media exposure. The level of active or non-active reaction such as excited-bores (arousal) of sympathetic nervous system (SNS) interprets the skin's ability to conduct electricity under the skin. The potential resulted from SNS is the tonic skin conductance level (SCL) and rapid phasic components or skin conductance responses (SCR). The SCR reacts to the arousal and differs from one person to another, thus making it a better choice for this study. This article described the various aspects of human emotions recognition design framework, including the emotions elicitation strategy, the type of stimuli, the targeted emotions, the human subject criteria, the suitable wearable sensors, experimental methods and signal processing activities.

2. Methods of Framework Design

The experimental methodology reflects the setup procedures for in-lab experiment. The experiment framework was prepared based on the four general steps in the signal processing activity as shown in Fig. 1.

Fig. 1. Overall process flow.

A wireless wearable biomedical equipment chosen for the data acquisition step to minimize the equipment complexity and to provide a comfortable environment to the research participants. Raw data were wirelessly streamed to a computer via Bluetooth communication. Raw signals contained noises from motion artefacts, electrodes and power interference. Typically, a collected raw signal contains noises involved very low frequency and very high frequency. A filtering process usually employed to reject the noises in a signal. Once the desired frequencies identified in a signal, filtering approach will be selected. The output of filtering process will maintain all the signal information and eliminate the unwanted frequencies. In feature extraction stage, the information were extracted from the filtered signal to describe the meaningful features of the recorded signals. Meaningful result of the feature extraction trained or processed using a computational analysis tool in the classification stage. At the end, the signal will be categorized into different emotion classes.

Two dimensional Valence-Arousal (V-A) emotion model reported by Lang [7] proposed that each emotion activated in unique and specific neural pathways. The Valence-Arousal model (V-A) can used to identify the type of emotions where Valence is how negative or positive such as happy-unhappy and Arousal is how calming or exciting. Negative emotions cause distractions to the driver and can seriously modulate attention and influence decision making abilities. Negative emotions that fall under negative (unpleasant) Valence and activated Arousal model were found more critical for survival [8].

The scope of this work focused on the data collection of local Malaysian populations. The measurement and recording of ECG and SCR signals were used in-lab equipment. The number of electrodes and sensors were limited by device specifications. Six emotions were investigated in the stimuli process which consisted of negative and positive emotion that included happy, sad, angry, fear and disgust, and also individual neutral state as control. A general research framework design of this work is shown in Fig. 2.

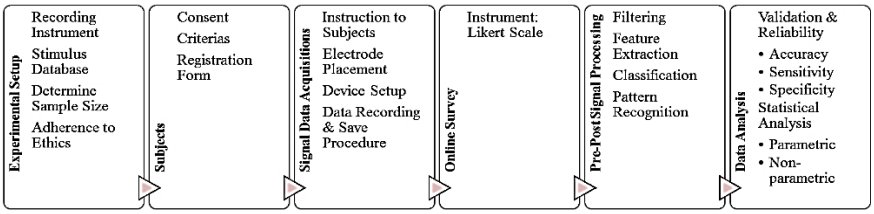

Fig. 2. Framework of human emotion recognition.

2.1. Stimuli set and emotion elicitation protocols

Spontaneous feelings are reflected by physiological changes. The James-Lange's theory was applied in this work where the affective stimulus occurred first, then peripherally physiological responses appeared such as rapid heartbeat, constriction of blood vessels and followed by perceived emotion [9].

Specific contents of the stimuli database were used to encourage the respondent in giving the respective responses. This process is called emotion elicitation. In the emotions stimuli database, researchers used images [10], sounds [11], audio visual [12], driving simulator [13], driving a car equipped with sensors in a real world environment [14], flash-cards, and set of selected words to evoke the emotions. Various emotion elicitation methods have been reported such as visual using pictures of International Affective Picture System Pictures (IAPS) and International Affective Digital Sounds (IADS) [10]. The assessments on the effects of music's valence on driving behaviour had been reported by Pêcher et al. [15] showed that happy music distracted drivers the most by referring to the mean speed and driving control and sad music directly influenced the driver to drive slowly and kept their vehicle in the lane. Words used in the roadside billboards also showed that drivers had lower mean speeds when emotional words were displayed, compared to neutral words [8]. In this work, an image stimuli was used because images have been globally accepted in various reported studies to elicit the emotion [16]. Video and audio visual (dynamic stimuli) had also developed in the stimuli process as video clips induced stronger positive and negative effects than music clips [17].

Prior to the stimulus database development, a survey was conducted to confirm the efficacy of the selected stimuli. Ten digital images and ten video-audio clips of each emotion were carefully chosen from public domain. Forty-five respondents participated in the survey and gave their feedback of the display stimuli that they believe had effectively evoked their emotions. These stimuli were closely related to the participant's culture, native language, and geography. Five emotions were investigated in the preliminary study, including happiness, sadness, anger, disgust, and fear. From this survey, we confirmed that the proposed stimuli achieved high efficacies of the studied emotions compared with an international stimuli database; IAPS and FlimStim [12]. In this survey, the efficacy of proposed stimuli used to evoke happy, sadness, anger, disgust and fear emotions obtained 72.9%, 77.17%, 80.34%, 76.87% and 67.36% respectively. Conversely, the efficacies of IAPS database used in the process stimuli obtained 62.88%, 57.08%, 59.8%, 45.07% and 65.69% of the same set of emotions. This result approved the hypothesis that the proposed stimuli suits the participant's population geography and cultural aspects of the study. Moreover, high image and video clip resolution are crucial for viewing satisfactions. The protocol used for data acquisition in this framework is shown in Fig. 3.

The stimuli session was segmented to three sections; picture display, video clip with sound and video clip without sound. Each section consisted of five emotional states, mainly under the negative arousal and valence; happy, sad, anger, disgust and fear. The arrangement of the emotions in the display stimuli was crucial in order to obtain high stimuli effectiveness. In this framework, each picture display contained five pictures and was displayed for 4 seconds each [18]. The cooling period between different emotions was twenty seconds and a black screen will be

displayed during this period [19]. One video clip used in video and audio visual stimuli. Three minutes of neutral state was recorded prior to the emotion elicitation process. Neutral images such as beautiful sceneries and meditation sounds containing water and soft music were played during the neutral stimulus. The entire session took about 40 minutes, including briefing, filling out of forms and consent, experimental setup, electrode placement, and signal recording.

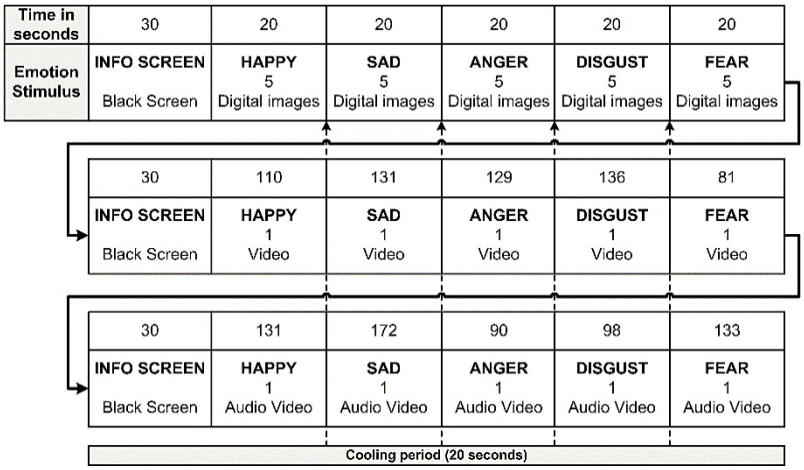

Fig. 3. Image and video clips stimuli protocol.

Dry and wet snap cloth electrodes were connected to a device through wires for recording. A commercial wireless bio-device (2.4 GHz IEEE protocol) used in this research was the BioRadio 150 Unit by Great Lakes NeuroTechnologies [20]. The collected signals were then exported into a .csv format and processed in Matlab processing tools. The sampling frequency (f_s) was 960 *Hz* with 12 bit resolution to ensure sufficient bandwidth (11520 bps) to record multi-psychophysiological signals simultaneously. As far as the storage was concerned, the file size of the recorded data can be estimated by considering the system bits, the sampling rate (total number of samples per second), recording time (in second) and total number of channels used to record the signal.

A questionnaire was created in Google Form and filled up by the respondent after the stimuli session. The survey questions were rated by using the 10 points Likert Scale [21]. This survey was conducted to investigate the efficacy of the proposed stimuli database used in the experiment.

2.2. Subject populations, instructions to the subjects and consent

Participants must be of Malaysian origin, aged between 18 to 60 years old, have a Malaysia's driving license, and have a real-world driving experience. Participants were briefed on the overall objective of the projects, data acquisition process, instructed to not tense up and stay relaxed for artefacts prevention [22]. All electronic devices such as mobile phones, WiFi connections, mobile modems, tablets and laptops were powered off. Watches and metal accessories were

removed from participant's body. All subjects must satisfy the criteria prior to the signal data recording.

The potential subjects will be excluded from participating in the study if they met any of the criteria such as a subject who refused to give consent, subject who had a problem of normal or corrected normal vision and hearing difficulties. Subjects were also informed that emolument will be granted for their participation. The participants were seated about 60 inches away from the 40 inch LCD TV and were rested for three minutes after being equipped with the sensors for signals calibration. The electrodes were firmly placed on the skin for a stable current flow and given sufficient time for the conductivity gel to penetrate into the hairy skin [23]. Experiment protocol was approved by Ethics Committee Involving Human Subjects, Universiti Putra Malaysia with accordance to Declaration of Helsinki.

2.3. Signal data acquisition

The crucial part of pre-signal processing is to determine the sampling rate (Hz), size of bit, and resolution [24]. In this work, a radio frequency (RF) type of device was used for recording and multiple channels were utilized, hence the sampling rate was increased to 960 Hz to ensure a sufficient bandwidth of data transmission. Some signal smoothing algorithms require signal down-sampling; however, the original input signal is restricted in high sample rates to avoid altered signals and information lost [21]. In this work, signals were recorded using cloth snap wet electrode and dry electrode to capture ECG and SCR signals. Adhesive electrodes provide cleaner, consistent and accurate output signal [25]. Skin surface wet electrodes were used in this work to measure the ECG potentials from the surface of the skin as shown in Fig. 4.

Fig. 4(a). Snap wire.

Fig. 4(b). Selected disposal electrode.

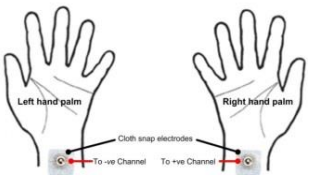

Fig. 4(c). Electrode placements of ECG.

The EDA sensor consisted of two snap electrodes (use of differential amplifier) that were placed on the tips of the fingers, typically the index and

middle fingers as depicted in Fig. 5. EDA measures the skin conductance proportional to sweat secretion; thus, EDA is usually measured at the palmar sites of hands or feet, which contain the highest sweat gland density (>2000/cm^2) [26].

Fig. 5(a). Non-disposal electrodes.

Fig. 5(b). SCR electrode placement.

2.4. Characteristic of ECG and SCR signals

The pre-processing stage is extremely important in preparing the cleanest, undistorted and retained information in the signal data. In many cases, the actual signal frequencies are much smaller than the noise frequency. Power source (AC power line interference) is one type of noise when any device is plugged into the power point usually centred at 50 Hz. Baseline wander is usually detected due to gross movements, mechanical strain on the electrode wires and improper electrode gel on the skin. It causes the signal's origin to drift away from the baseline point. Finite impulse response (FIR) and infinite impulse response (IIR) are reported to efficiently remove the baseline noises [27]. FIR filter is developed digitally, where the desired magnitude is implemented in discrete time domain and produced stable output, however FIR filter needs more filter order [28]. IIR used to operate under a non-linear operation but can lead to instability and difficult to control [29]. Noises caused by muscle movement (EMG) frequently infuse artefacts at very high frequency (>100 Hz).

Figures 6 and 7 summarize the important frequency components of ECG and SCR respectively. The normal characterization of ECG is usually explained in wave types and wave intervals [30]. Wave types known as P, Q, R, S and T represent the activities of the atrial and ventricular parts in the heart in specific time duration (in unit seconds) and amplitude (in unit mV). ECG signal's interval is referring to the time taken from one wave type to the other ECG wave types. In medical practice, these ECG characteristics were used to investigate any abnormalities in the recorded signals. Usually, prolong or shortened intervals detected in the signal help the medical practitioner to understand the condition of the patient's heart.

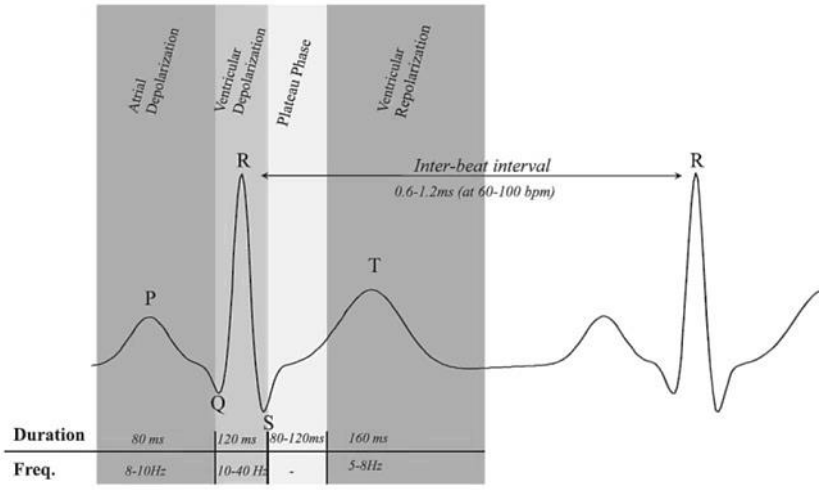

Fig. 6. Normal heartbeat of ECG signal [6].

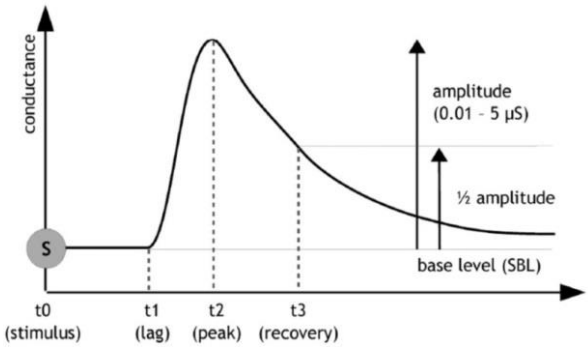

Fig. 7. Normal SCR signal [31].

Normal characterization of SCR waves are divided into five main segments; latency of the response (lagging phase) that occurs between 1 to 3 seconds, the signal's rise time from the latency up to its response peak recorded between 0.5 to 3.5 seconds, the amplitude of the response peak at 0.01 to 5 µSiemen which occurs between 1.5 to 6.5 seconds, the half time value of the recovery time which occurs between 1 to 10 seconds and finally, the signal threshold identified from the signal's amplitude 0.01 to 0.05 µSiemen [21, 31, 32].

The frequency contents of ECG and SCR signals frequently represented by using Fast Fourier Transforms (FFT). Fast Fourier Transforms frequently suffers losses of information in the time domain and gives only spectral information in the frequency domain and vice versa. Short Time Fourier Transform (STFT) represents the signal in both time and frequency domains using the moving window function to overcome the information loss problem in FFT. Then, the

performance evaluation of filtered signals can be conducted by performing the Signal-to-Noise Ratio (SNR) tests for both before and after filtering [33].

The calculated features of filtered signals can be up to hundreds of features from various feature domains which will then be identified for the emotion-relevant features by using the feature selection method. Basic calculation of the total number of features for classification can be explained as in Eq. (1):

$$\text{(number of emotions evoked)} \times \text{(number of stimuli segments)} \times \text{(number of selected features)} \times \text{(number of subjects)} \times \text{(number of channels)} \quad (1)$$

where the number of emotions refers to the studied emotions (in this work they were happy, anger, sad, fear and disgust); the number of stimuli segments refers to the selected stimuli method such as image, sound, video clips; number of subjects refers to the total number of respondent; number of selected features are the syntactical and statistical features that can be extracted from the denoised signal such as the signal's mean, median, standard deviation, variance, number of peak counted and sum of peak amplitude in a examined window, power spectral, entropy and others; and the number of channels is the total number of channels used to record the biological signals [34].

Signal segmentation is implemented before meaningful features can be extracted due to long recording time. Some techniques presented in previous works included the fixed size segmentation technique, overlap windowing technique to preserve signal information [35], ECG signal beat detection [36] and event related routines [21]. Usually too many features are not useful for producing a desired learning result. A reduced feature representation will complete the tasks faster; hence it makes the system computationally less-expensive. The pre-processing general steps of this work summarized as below:

Initialize

1. Define stimulus session; remove signal calibration segment from raw signal
2. Filtering;
   - Frequency spectrum analysis
   - Identify desired signal and noise frequencies
   - Select filter design
   - Define cut-off frequency
   - Confirm output using SNR and frequency spectrum analysis
3. Filtered signal segmentation; according to display stimulus
4. Defined sampling rate for each segment; (stimulus duration × sampling rate) for time domain analysis
5. Feature extractions and selections
   - ECG: R peak and RR interval (mean, median, mode, variance, standard deviation, interquartile, skewness, and kurtosis values also heart rate in beat per minute)
   - SCR: Total number of peak and sum of peak amplitude (mean, median, mode, variance, standard deviation, interquartile, skewness, and kurtosis values)

Normalization (feature scaling) is a method used to standardize the range of data features. Normalized features usually in the limit of magnitude such as scaled -1 to 1, 0 to 1 or (data/maximum data) where target range is depends on the nature of the processed data. From the mix distribution of the calculated features, SVM classifier was selected to obtain the classification accuracy of the studied emotional states. For non-separable classes, SVM used a soft margin where the hyperplane separates data points at its best effort. The selected model of SVM classifier involved a routine to choose the best hyperplane based on the tested radial basis function.

3. Pilot study observations and early findings

A pilot study was conducted to authenticate the designed framework. A complete set of stimuli database as shown in Fig. 3 has been utilized. Early findings and observations are vital in order to get a clear picture of any issue or potential obstacles that may occur during the full scale experiments. Hence, experiment procedures and protocols can be revised. In the pilot experiments, twenty three volunteers (age 23 to 36, 8 males and 15 females) were participated. Four subjects were excluded due to high movement artefact and truncated signals. The collected signals were processed by using Matlab and ran in Intel CORE i5 processor. The observations and findings in the pilot experiment are deduced in the following subsections.

3.1. Stimuli database and protocols

There were several issues in the stimuli database that had been identified. Respective improvements will be introduced in the future experiment to ensure high emotion recognitions classifications accuracy, sensitivity and specificity. The observations and findings in the pilot experiment were deduced in the following section. Survey results are depicted in Fig. 8.

From the participant's feedback, the video clips exhibited an excellent efficacy of 82.6% to 95.7% except the clip used to elicit the fear emotion. The analyzed data from the survey conducted after the emotion elicitation session showed that digital images stimuli used to evoke happy emotions yielded the highest efficacy of 82.6%. However sad, anger and fear elicited by using the digital image stimuli exhibited lower efficacy at 65.2%, 78.2% and 69.6% respectively. Low efficacy percentage of the used stimulus will need to be revised and replaced with more relevant pictures or clips in order to evoke the emotions more effectively.

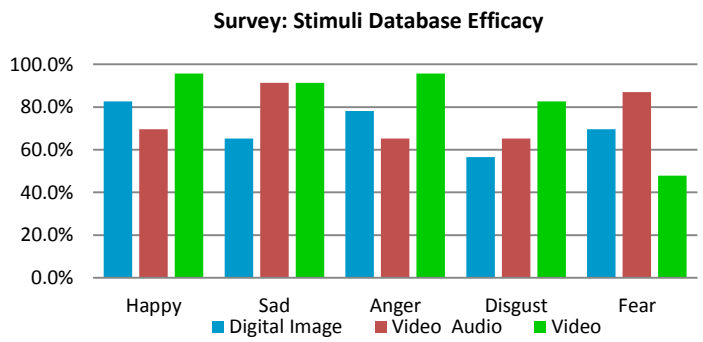

Fig. 8. Effectiveness of the stimulus obtained from survey method.

3.2. Experimental setup and signal data acquisitions

Our device operated wirelessly to transmit the collected signals to the computer. We found that the AC source from a computer or laptop changed the mean amplitude of the signal to about 8.46% when it was placed close to the device unit. The presence of WiFi connections including mobile modem internet, mobile phone or tablet will also interfere and change the collected signal at about 9.14% compared to when all these electronic devices were shut down. However, collecting multi-signals simultaneously showed insignificant impact to all the recorded signals.

Data lost or signal drop is a common issue in wireless technology. We found signal drop of the BioRadio device did not affect the ECG signal transmission, as the ECG was able to return to its baseline instantly. However, the signal drop caused a major problem to the SCR signal due to its tiny output voltage (µSiemen). This issue required some adjustments to the USB serial port driver to optimize the system performance and make the BioCapture software find the USB Receiver (which was connected to the computer over the USB port) more consistently. The latency timer of the timer that read and wrote timeouts value in the computer's device manager COM port number must be changed to the smallest value (1 millisecond) to avoid severe SCR data lost.

Cloth surface electrode contact was diminished when sweating occurs. Considering the stimuli process took nearly 30 minutes continuously and cannot be stopped or repeated, room temperature played an important aspect in order to ensure sufficient skin to electrode contact. For the best ECG signal display during high levels of activity, the shielded lead electrode cable and high performance foam electrodes were used in place of the "button snap" electrode cables or cloth surface electrodes.

We observed that the first 5 to 10 minutes rest time was crucial in order to ensure that the body systems were in a relaxed mode. Calibration period took up to three to five minutes to ensure that the conductive gel or electrode gel had penetrated fully into the skin especially the ECG signal. During this period, the investigator should look into the recorded signals to identify any unusual signal shapes due to electrode placement or device pre-setting issue.

A differential amplifier (consisting of positive and negative input voltages) was used to perform subtraction between the two channels of ECG signals. The positive channel point electrode of ECG should be placed on the right side of the wrist/chest and the negative channel point electrode placed on the left side of the body for better signal acquisitions. Moreover, when the same device was used to collect multi-physiological signals simultaneously, the suggested ground point by the manufacturer was on the middle forehead (namely F_{PZ} point which referring to electroencephalography signal data), dangling cables were twisted, bundled together, and taped to the skin with a medical tape.

3.3. Signal processing

The valence states were found to be better explained by the heart rate (HR) features that obtained from ECG signals [18]. The ECG beat per minute (BPM) values can be calculated by referring to the R to R interval. The R peak and its peak location were obtained using find local maxima algorithm in Matlab where

the indices at which the peaks occur were returned. Then heart rate variability (HRV) was calculated using the R peak and location information representing the average of heart rate changes from beat to beat (R to R intervals) in a minute time frame. The frequency component of ECG are ranged between 0.05 Hz to 150 Hz for diagnostic and 0.5 Hz to 40 Hz for monitoring.

In this work, a bandpass signal filtering technique using 4th order bandpass Butterworth type filter with cut off frequency 100 Hz was used to eliminate muscle artifacts and 0.5 Hz to discard low frequency components mainly due to motion artifacts and respirations variation. The power line interference was eliminated from ECG by using the Butterworth bandstop filter (cutoff frequency 48-52 Hz). The Butterworth filters sufficiently removed the noises because of the flat passbands, acceptable roll of time and wide transition bands. Baseline noises were eliminated by using detrending data technique. The segmentation of the filtered signals was performed according to the emotion stimulus depicted in Fig. 3.

In the classification stage, five features were selected that including the total number of R peak in each evaluated window, mean and variance values of R peak also the mean and variance values of R-R interval in the stimulated segments. Each emotion was cross validated with the individual's neutral state. A 2-D Gaussian mixture model with holdout option of the support vector machine (SVM) technique was chosen for the classification stage. Table 1 summarizes the early finding of human emotion recognition using ECG signal using the designed protocols.

Table 1. ECG emotion classification results.

Stimulus	Emotions	Accuracy	Sensitivity	Specificity
Digital images	Happy	59.2	66.9	48.4
	Sad	60.8	69.2	48.3
	Anger	56.9	65.1	44.7
	Disgust	62.4	76.2	40.9
	Fear	62.4	66.3	56.1
Audio Visual	Happy	69.2	73.5	63.2
	Sad	71.1	70.9	71.3
	Anger	72.7	80.4	62.9
	Disgust	73.8	77.4	69.4
	Fear	72.8	78.4	61.5
Audio	Happy	59.4	67.9	66.8
	Sad	72.8	80.2	65.8
	Anger	69.1	73.9	60.1
	Disgust	68.9	68.1	70.1
	Fear	71.6	71.6	71.7

In ECG signals, the used of digital images in emotion elicitation produced the lowest emotion classification rate compared with the video and audio-visual methods. The specificity and very low sensitivity results had emphasized on ineffectiveness of the digital images used in this study. The selected clips used in the video and audio-video stimulus were found more effective to evoke the emotions with accuracy rate that higher than 68%. The specificity and specificity results also shown that the classifier had sufficiently identified the studied emotions and neutral emotional state. From these findings, we can conclude that

the rated video and audio visual display stimuli had successfully evoked the emotions by using the ECG signal.

In SCR, two extracted features namely number of peak and the sum of peak amplitude counted in each 5 seconds of window frame. A bandpass Butterworth filter with cutoff frequencies of 0.5-2 Hz was sufficiently removed the power line noise, low frequency drift and artefact removal. Figure 9 represents one of the selected SCR signal before and after the pre-processing stage.

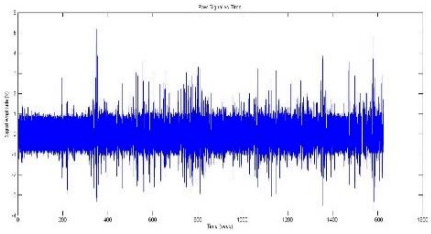

Fig. 9(a). Raw SCR signal.

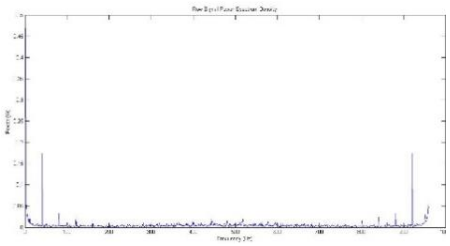

Fig. 9(b). Raw SCR signal frequency spectrum.

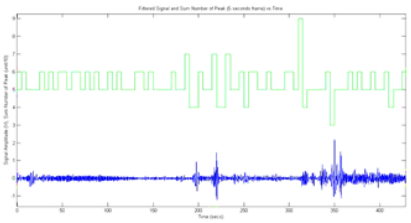

Fig. 9(c). Filtered SCR signal and total number of peaks.

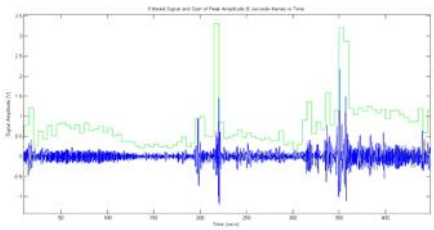

Fig. 9(d). Filtered SCR signal and sum of peak amplitude.

The initial finding of emotions classification using SCR signal summarized in Table 2. The detected happy, sad, anger, disgust and fear emotions evoked by using selected video clips have shown the highest classifications accuracy among the other stimulus methods. The accuracy was ranged from 67.1% to 72.1%. In contrast, the stimulus method that used digital images produced 43.1% to 62.5% emotion classification accuracy while video-audio clips shown 40.2% to 53.2% accuracy. We discovered two emotions; happy and anger emotions evoked by using digital images, video-audio and video clips have shown similar patterns in accuracy rate and in the stimulus effectiveness which obtained from survey stated earlier.

In this experiment, we also learned skin conductance level contains both AC and DC components. The phasic waveform (SCR) can be separated from SCL signal by filtering SCL signal as shown in Figs. 10 and 11. Typical SCL's baseline is not uniform and each person has a different SCL with tonic levels ranging from 10 to 50 µSiemen. Tonic skin conductance levels are varying over time in every individual, depending on individual psychological state and autonomic regulation. This finding shows that SCR signal is more suitable to use in this work as it contains specific and stable signals characteristics compare to SCL signal.

Table 2. SCR emotion classification results.

Stimulus	Emotions	Accuracy	Sensitivity	Specificity
Digital images	Happy	43.1	57.6	68.4
	Sad	50.0	65.0	69.7
	Anger	62.5	73.7	65.5
	Disgust	61.1	69.7	76.8
	Fear	59.7	65.0	62.4
Audio Video	Happy	40.2	59.3	68.5
	Sad	53.2	55.9	61.9
	Anger	50.3	70.7	55.3
	Disgust	50.0	65.7	68.2
	Fear	48.6	73.4	57.0
Video	Happy	70.2	57.7	74.0
	Sad	67.1	60.0	65.9
	Anger	70.0	60.0	65.9
	Disgust	72.1	59.0	74.0
	Fear	70.9	58.1	74.9

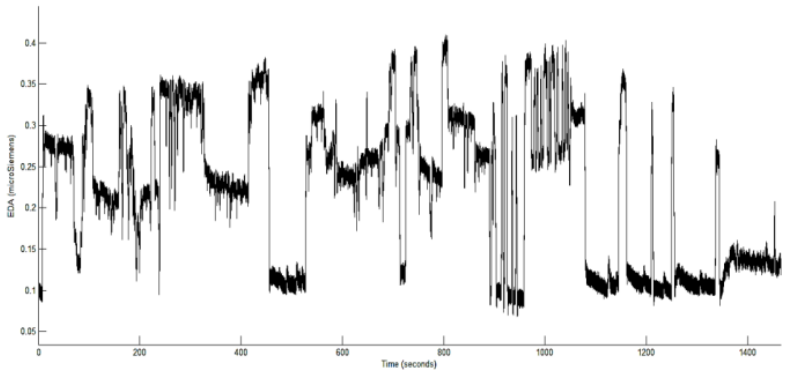

Fig. 10. Raw SCL signal.

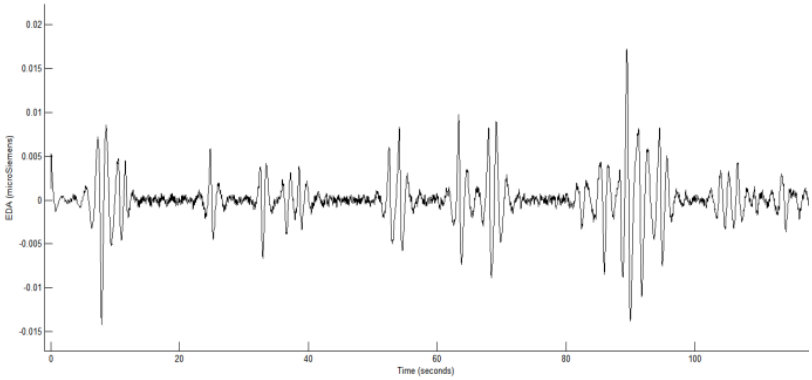

Fig. 11. SCR component in SCL. Baseline shown SCR characteristic after filtered SCL signal.

4. Discussions

This article presented a systematic review of design frameworks to recognize human emotions using ECG and SCR signals. The aim is to investigate the correlation of multi-psychophysiological signals response to the emotions change. New stimuli protocol considered the local culture and display stimulus were developed to evoke the emotions. We can deduce that the video and audio visual stimulus method have showed a great potential to detect the emotions using ECG and SCR signals. We found that multiple emotions stimulated continuously affected the overall emotion elicitations effectiveness. In the future, system modelling will be focusing on a negative and positive emotions in order to gain the knowledge of the specific ECG and SCR signal's pattern during angry and happy emotions.

Low classifications accuracy in this pilot works were expected due to the small sample size, which caused the classifier has very limited data to perform the train and test efficiently. Moreover the effectiveness of the neutral state using neutral images and soothing music may insufficient to define emotional baseline. In the future work, a resting state might be employed where biological signals recorded during the subject is not performing an explicit task.

In order to be convinced about the success of the proposed method, the hypothesis testing by using two population means can be used to determine the sufficient number of the sample population size of emotion recognition work in the future. Two-tailed test consisted of $\mu_1 \neq \mu_2$ where $\mu_1 = 0$ referring to no-emotional affect detected and $\mu_2 > 0$ referring to the sample mean based on the mean value that had been reported in previous works. The sample size can be determined using Eq. (2):

$$n = \frac{2\sigma^2 \left[z_{1-\alpha/2} + z_{1-\beta} \right]^2}{(\mu_1 - \mu_2)^2} \tag{2}$$

where the sample population size is representing the n, μ_1 is the mean of the first population, μ_2 is the mean of the second population, σ^2 is the sample variance calculated from the mean, sample standard deviation is σ, $z_{1-\alpha/2}$ is 1.96 and $z_{1-\beta}$ is 1.64(referring to the z table of two sided test value), with the assumption of significance level at 95% and confidence interval $\alpha = 0.05$ and $\beta = 0.1$. The mean of the sample population μ_2 is referring to the successful happy, anger, sad, disgust and fear emotions recognition obtained by using ECG and SCR signals.

Moreover, different classifiers other than SVM will be comparatively tested on the collected data. Also, different feature extraction methods and the motivation behind the selected features will also be investigated. The improvements in the stimulus protocol and the stimuli display used in the video session are needed to evoke the specific emotions more effectively. Since each individual's emotion state was categorized by referring to their neutral state, the recorded neutral stimuli should be longer than all image and video clip displays used in the stimuli database in order to get better chances of a good neutral state.

In the future, the impact of external and internal temperature to the selected ECG and SCR electrodes and also sensors placed on the skin or fingers will be investigated due to the concern of high humidity in this region. High humidity

causes sweat and will not evaporate into the air and we feel much hotter than the actual temperature.

5. Conclusions

A Proof-of-Concept of the designed and proposed framework of this work had shown that both ECG and SCR had successfully elicited positive and negative emotions. Three different stimuli techniques (images, video clips and video-audio clips) proposed had also produced an extensive number of information for a deeper descriptive analysis in the future and had anticipated a great potential for our framework to succeed. The findings of this research were expected to contribute in the future automated real-time affective recognition systems. In this era, the integration of technology aids including intelligence facilities of HMI (Human Machine Interface) and BCI (Brain Computer Interface) are built in the vehicles. This work is a step towards a complete emotion recognition system effort to reduce the emotional behaviour of the driver that led to road accidents. Enhanced system to analyse the physical and psychological state of the individual will eliminate the driving risk factor, specifically for the rescue team and commercial vehicle drivers. The development of affordable, reliable, and noncomplex sensor systems in vehicles remains crucial in accident prevention research area. The findings of this work can contribute to road safety and automotive industry advancement. The driving risk factors can be eliminated through enhancement of the systems utilizing psychological and physiological aspects.

Acknowledgment

This work was financially supported by Ministry of Science Technology and Innovation, Malaysia (Grant No. 01-01-02-SF1061). Funders were not involved in the conduct of the research.

References

1. Ahmad Noor Syukri, Z.A.; Siti Atiqah, M.F.; Fauziana, M.; and Abdul Rahmat, A.M. (2012). MIROS crash investigation and reconstruction annual statistical report 2007-2010. *Annual Statistical Research Report No. MRR 05/2012,* Malaysian Institute of Road Safety Research, Kuala Lumpur.
2. Berdoulat, E.; Vavassori, D.; and Sastre, M.T.M. (2013). Driving anger, emotional and instrumental aggressiveness, and impulsiveness in the prediction of aggressive and transgressive driving. *Accident Analysis & Prevention*, 50, 758-67.
3. Schwebel, D.C.; Severson, J.; Ball, K.K.; and Rizzo, M. (2006). Individual difference factors in risky driving: The roles of anger/hostility, conscientiousness, and sensation-seeking. *Accident Analysis & Prevention*, 38(4), 801-810.
4. Hosseini, S.A.; and Khalilzadeh, M.A. (2010). Emotional stress recognition system using EEG and psychophysiological signals: Using new labelling process of EEG signals in emotional stress state. *Proceedings of IEEE Biomedical Engineering and Computer Science Conference*, 1-6.
5. Seoane, F.; Mohino-herranz, I.; Ferreira, J.; Alvarez, L.; Llerena, C.; and Gil-pita, R. (2014). Wearable biomedical measurement systems for assessment of

mental stress of combatants in real time. *Sensors*, 14(4), 7120-7141.

6. Agrafioti, F.; Hatzinakos, D.; Member, S.; and Anderson, A.K. (2012). ECG pattern analysis for emotion detection. *IEEE Transactions on Affective Computing*, 3(1), 102-115.
7. Bradley, P.J.; and Lang, M.M. (1994). Measuring emotion: The selff-assessment manikin and the semantic differential. *Journal of Behavioral Therapy and Experimental Psychiatry*, 25(1), 49-59.
8. Chan, M.; and Singhal, A. (2013). The emotional side of cognitive distraction: implications for road safety. *Accident Analysis & Prevention*, 50, 147-154.
9. Calvo R.; D'Mello, D. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, 1(1), 18-37.
10. Bradley, M.M.; and Lang, P.J. (2007). The International Affective Picture System (IAPS) in the study of emotion and attention. *Handbook of Emotion Elicitation and Assessment*, 29-46.
11. Selvaraj, J.; Murugappan, M.; Wan, K.; and Yaacob, S. (2013). Classification of emotional states from electrocardiogram signals: A non-linear approach based on hurst. *Biomedical Engineering Online*, 12(1), 44.
12. Schaefer, A.; Nils, F.; Sanchez, X.; and Philippot, P. (2010). Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognitive Emotions*, 24(7), 1-36.
13. Wulf, F.; Rimini-döring, M.; Arnon, M.; and Gauterin, F. (2014). Recommendations supporting situation awareness in partially automated driver assistance systems. *IEEE Transactions on Intelligent Transportation Systems*, 16(4), 2290-2296.
14. Vicente, F.; Huang, Z.; Xiong, X.; De Torre, F.; Zhang, W.; and Levi, D. (2015). Driver gaze tracking and eyes off the road detection system. *IEEE Transactions on Intelligent Transportation Systems*, 16(4), 2014-2027.
15. Pêcher, C.; Lemercier, C.; and Attenti, J.-M.C. (2009). Emotions drive attention: Effects on driver's behaviour. *Safety Science*, 47(9), 1254-1259.
16. Henderson, R.R.; Bradley, M.M.; and Lang, P.J. (2014). Modulation of the initial light reflex during affective picture viewing. *Psychophysiology*, 51(9), 815-818.
17. Lazar J.N.; and Pearlman-Avnion, S. (2014). Effect of affect induction method on emotional valence and arousal. *Psychology*, 5(7), 595-601.
18. Swangnetr, M.; and Kaber. (2012). Emotional state classification in patient-robot interaction using wavelet analysis and statistics-based feature selection. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 43(1), 1-13.
19. Bos, D.O. (2006). EEG-based emotion recognition: The influence of visual and auditory stimuli. *Capita Selecta*, 1-17.
20. Cleveland Medical Devices Incorporation. (2011). BioRadio® 150 User's Guide. Retrieved from https://glneurotech.com/bioradio/bioradio-150/.
21. Braithwaite, J.; Watson, D.; Robert, J.; and Mickey, R. (2013). A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology*, 49, 1017-1034.
22. Fatourechi, M.; Bashashati, A.; Ward, R.K.; and Birch, G.E. (2007). EMG

and EOG artifacts in brain computer interface systems: A survey. *Clininical Neurophysiology*, 118(3), 480-494.

23. Huang, Y.; Wu, C.; Wong, A.M.; Lin, B. (2015). Novel active comb-shaped dry electrode for EEG measurement in hairy site. *IEEE Transactions of Biomedical Engineering*, 62(1), 256-263.
24. Ghasemzadeh, H.; Ostadabbas, S.; Guenterberg, E.; and Pantelopoulos, A. (2013). Wireless medical-embedded systems: A review of signal-processing techniques for classification, *IEEE Sensors Journal*, 13(2), 423-437.
25. Jeon, T.; Kim, B.; Jeon, M.; and Lee, B. (2014). Implementation of a portable device for real-time ECG signal analysis. *Biomedical Engineering Online*, 13(160), 1-13.
26. Bornoiu, I.-V.; and Grigore O. (2013). A study about feature extraction for stress detection. *Proceedings of IEEE Advanced Topics in Electrical Engineering Symposium*, 1-4.
27. Bhogeshwar, S.S.; Soni, M.K.; and Bansal, D. (2014). Design of simulink model to denoise ECG signal using various IIR & FIR Filters. *Proceedings of IEEE International Reliability, Optimization and Information Technology Conference*, 477-483.
28. Shoeb A.; and Clifford, G. (2006). Wavelets; multiscale activity in physiological signals. *Proceedings of Biomedical Signal Image Processing Symposium*, 1-29.
29. Karthikeyan, P.; Murugappan, M.; and Yaacob, S. (2012). ECG signal denoising using wavelet thresholding techniques in human stress assessment. *International Journal of Electrical Engineerng Informatics.* 4(2), 306-319.
30. Seena V.; and Yomas, J. (2014). A review on feature extraction and denoising of ECG signal using wavelet. *Proceedings of IEEE Devices, Circuits and Systems Conference*, 1, 1-6.
31. Leiner, D.; Fahr, A.; and Früh, H. (2012). EDA positive change: A simple algorithm for electrodermal activity to measure general audience arousal during media exposure. *Communication Methods and Measures*, 6(4), 237-250.
32. Schmidt, S.; and Walach, H. (2000). Electrodermal activity (EDA)-state-of-the-art measurement and techniques for parapsychological purposes. *Journal of Parapsychology*, 64(2), 139-164.
33. Goldberger, A.L.; Amaral, L.A.N.; Glass, L.; Hausdorff, J.M.; Ivanov, P.C.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.-K.; and Stanley H.E. (2013). PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215-e220.
34. Vinhas, V.; Oliveira, E.C.; and Reis, L.P. (2010). BioStories-dynamic multimedia interfaces based on automatic real-time user emotion assessment. *Proceedings of International Information Systems Conference*, 21-29.
35. Uy, G.; Ngo, C.F.; Trogo, R.; Legaspi, R.; and Suarez, M. (2013). Correlation of stress inducers and physiological signals with continuous stress annotations. *Proceedings of Information and Communications Technology*, 7, 178-183.
36. Ge, D.; Srinivasan, N.; and Krishnan, S.M. (2002). Cardiac arrhythmia classification using autoregressive modeling. *Biomedical Engineering Online*, 1(1), 1-5.