34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Human Behavior State Profile Mapping Based on Recalibrated Speech Affective Space Model

N. Kamaruddin, and A. Wahab, *Member, IEEE*

Abstract — **People typically associate health with only physical health. However, health is also interconnected to mental and emotional health. People who are emotionally healthy are in control of their behaviors and experience better quality of life. Hence, understanding human behavior is very important in ensuring the complete understanding of one's holistic health. In this paper, we attempt to map human behavior state (HBS) profiles onto recalibrated speech affective space model (rSASM). Such an approach is derived from hypotheses that: 1) Behavior is influenced by emotion, 2) Emotion can be quantified through speech, 3) Emotion is dynamic and changes over time and 4) the emotion conveyance is conditioned by culture. Empirical results illustrated that the proposed approach can complement other types of behavior analysis in such a way that it offers more explanatory components from the perspective of emotion primitives (valence and arousal). Four different driving HBS; namely: distracted, laughing, sleepy and normal are profiled onto the rSASM to visualize the correlation between HBS and emotion. This approach can be incorporated in the future behavior analysis to envisage better performance.**

I. Introduction

A good quality of life can be achieved when an individual physiological and psychological need were fulfilled. In fact aligning ones behavior to the norm can lead to self-satisfaction. Human behavior normally reflects the way a person processes the information from the environment (stimuli) and reacts to it (feedback). It needs to be executed in such a way that it is acceptable by the community at large. On the contrary, an abnormal behavior can be misconstrued as unhealthy behavior resulting in rejection by the community or group. Behavior also reflects both the conscious or subconscious emotion information communicated from the speaker to the audience. Thus understanding and recalibrating such a behavior can help one to lead a more normal life in a community. Based on this observation, there are a strong correlation between behavior and emotion, which will be investigated in this paper in order to profile the human behavior state (HBS).

Many researchers have adopted the discrete-class classification system to recognize emotion [1, 2]. This classification system referred to a selection of limited number of emotions to form the subset of emotion list. The most popular list is proposed by Cornelius [3, 4] which consists of emotion: anger, happiness, sadness, fear, disgust and surprised. Such an approach is employed to simplify the emotion classification task. A similar approach can also be adopted to analyze HBS. A small number of HBS; for instance, distracted or normal behavior states can be observed and discriminated. However, the task to get appropriate data that represents the actual real-time HBS posed a challenge. The naturalness of the data will be compromised if the data collected is acted or induced. Such a static approach also does not adequately explain the HBS since it can be erratic and impulsive. Thus, it is vital to have a real-time HBS data that allows the subjects to behave without any inhibition.

In addition, many researchers tend to focus on a small number of HBS class since specific data for a particular HBS is difficult to collect. Such an approach may be successful if the number of subject is limited and the subjects are conditioned in a same cultural influence [5]. This is because the way one behaves may be different from the others although they are given a similar set of scenario (in this case we assumed that people in a similar culture react similarly). Thus, a recalibrated speech affective space model (rSASM) derived from different cultural-influenced speech emotion data is proposed as a template to map the selected HBS. The proposed approach is based on the hypotheses that 1) Behavior is the manifestation of emotion, 2) Emotion is dynamic and changes over time, 3) Emotion can be empirically measured using speech and 4) The emotion conveyance and perception is conditioned by culture. Such profiling allows the HBS to be analyzed from the perspective of emotion primitives; namely: valence and arousal. Valence refers to the effect of emotion ranging from positive (pleasure) to negative (displeasure) effect whereas arousal describes about the activation level of emotion ranging from active to passive. These attributes function as constituents that act as a fully complementary description of the emotion. The usage of emotion dimensions such as positive-negative continuum would be descriptively relevant to a functionalist psychologist position insofar as the hedonic tone of the emotional experience is a notable marker and influenced subsequent behavior [3, 6, 7].

II. Data Corpus, Features Extraction and Classifier

A. Data Corpus

There are four corpora used in this experiment. It can be separated as the training dataset and the testing dataset. The training dataset is the used to derive the culturally-influenced

Norhaslinda Kamaruddin is a lecturer with the Faculty of Computer and Mathematical Sciences, MARA University of Technology (UiTM), 40400 Shah Alam, Selangor, MALAYSIA (phone: 603-5521 1130; fax: 603-5543 5502; e-mail: norhaslinda@tmsk.uitm.edu.my).

Abdul Wahab is now a Professor with the Faculty of Information and Communication Technology, Department of Computer Science, International Islamic University Malaysia (IIUM), 53100 Jalan Gombak, Kuala Lumpur, MALAYSIA (e-mail: abdulwahab@iium.edu.my).

978-1-4577-1787-1/12/$26.00 ©2012 IEEE

rSASM. Three different dataset of NTU_American [5], Berlin [8] and NTU_Asian [5] datasets are employed to represent the American, European and Asian cultures respectively. Comprehensive description of the data can be found in [5]. In this work, four emotions of anger, happiness, sadness and neutral (acting as emotionless state) are selected to exhibit the different emotion primitives' values as presented in Table 1. Emotion primitives' values of the emotions in Table 1 are used as a training set to generate the rSASM. Such values are derived from the psychologists' agreement that emotion can be represented using the characteristic of emotion [6, 7, 11]. Then, testing dataset values will be mapped onto the generated rSASM so that the correlation between HBS and emotion can be visualized.

TABLE I. INTENDED OUTPUT FOR DIFFERENT EMOTIONS

Emotion	Valence Value	Arousal value	Quadrant
Anger	-1	+1	2
Happiness	+1	+1	1
Sadness	-1	-1	3
Neutral	0	0	Center
Calm	+1	-1	4

Neutral is included in the analysis as it illustrates the origin of the newly recalibrated axis. Although psychologists claim that neutral is located at the (0,0) coordinate of the affective space model [7], Kamaruddin et al. reported in their recent study that the neutral coordinate is shifted due to the cultural effect [5]. Hence, neutral will be adopted as an origin axes for the proposed rSASM. With an assumption that emotion is universal, we hypothesize that similar emotion from different cultures share similar accoustical characteristics. Such inference justified the combination of the different corpora as the heterogeneous training data. Comparison between homogeneous training dataset (only one culture dataset) and heterogeneous training dataset (combination of two or three cultures) is needed to show the effect of inter-cultural in HBS.

The Real time Speech Driving (RtSD) corpus is recorded to act as a testing dataset under differing HBS [15]. It represents four real-time driving HBS; namely: distracted (talking on the mobile phone), laughing, feeling sleepy and driver having normal conversation. This dataset is collected from real-time natural speeches of the drivers while they are driving a motor vehicle. Such an approach allows freely expressed emotion to be captured while the drivers concentrate on controlling the vehicle. Detail explanation is provided in [9, 10, 15]. Subsequently, the methodology of this data collection is based on fair assumption that normal and distracted driver would behave differently and it is reflected prominently by their driving behavior.

The distracted HBS is selected based on the assumption that the human attention will be diverted because he/she had to multitask and divide the attention between two tasks and providing appropriate response to the caller on a cell phone. The laughed HBS is included because laughter is often used as a cue to manifest happiness for speech-only situations. Such HBS inclusion is suitable for positive valence-arousal analysis as happiness is located in the first quadrant of rSASM. Sleepy HBS is incorporated in the analysis to investigate the discrepancy of the Valence and Arousal values of the 'pure' sleepy and sleepy experienced during driving exercise. Russell in his paper [11] described sleepy as a state that has positive valence and negative arousal values together with calmness. However, the driver may not be experiencing positive effect when feeling sleeping especially during driving. This is because he/she needs to maintain his level of arousal by forcing himself/herself to stay awake hence resulting in a negative effect of the Valence. Such a scenario is true in most cases especially when you have to keep yourself awake in a class or meeting. This negative Valence values correspond to the negative emotion such as frustration or irritation. In addition, although sleepy is not considered as one of the Cornelius basic emotion [4], it is deemed substantial to include sleepy detection in the proposed HBS analysis because sleepy is known as one of the factor that contributes to unhealthy behavior especially in a class or meeting. Finally, the normal HBS is included to serve as a reference to the other HBS analyses. At this stage, the individual is experiencing neutral emotion at the initialization period prior to the commencement of the recording exercise. This HBS is important to gauge and measure the driver's emotion differences before and after he/she completed the given tasks and responses during the recording exercise. The speech data are then labelled by matching each experimental condition with its distinctive HBS label; denoted namely as: distracted, laughed, normal and sleepy HBS for profiling purpose.

B. Features Extraction

Features must be extracted from the pre-processed speech emotion data and transformed into an appropriate format for further processing. This is needed in order to effectively recognize the different emotions from the speech signal. To date, human auditory system outperforms current machine-based system for speech emotion recognition system. Human are known to be better at categorizing different emotion in almost automatic response. Motivated by this fact, Mel Frequency Cepstral Coefficient (MFCC) [12] features extraction method was adopted as it is based on the approximation of critical bands in the human auditory system. The performance of MFCC features extraction coupled with different classifiers to discriminate emotions are provided in [5, 9, 10, 13, 15]

C. Classifier

Adaptive Network-based Fuzzy Inference System (ANFIS) [14] is adopted as the classifier to compare the different dataset homogeneity performance in profiling the HBS on rSASM. ANFIS is a hybridization of neural network and fuzzy system that optimizes the parameters of the given fuzzy inference system by applying learning procedure to the training dataset. It uses the hybrid algorithm, which is the mixture of least mean square (LMS) error minimization to determine the consequent parameters and back-propagation

to learn the premise parameters. Further description of the ANFIS implementation is provided in [5, 13].

III. Recalibrated Speech Affective Space Model

The general approach of human behavior state (HBS) profiling based on recalibrated speech affective space model (rSASM) is illustrates in Fig. 1. Speech data for both training and testing datasets are transformed into its respective MFCC features. Then, ANFIS will train the training dataset features according to its emotion primitives' value. Table 1 summarized the intended output for the different emotions. Then, the rSASM will be constructed based on the output of the ANFIS training. Subsequently, the HBS profile will be mapped on this model to visualize the emotion primitives of the respective behavior states.

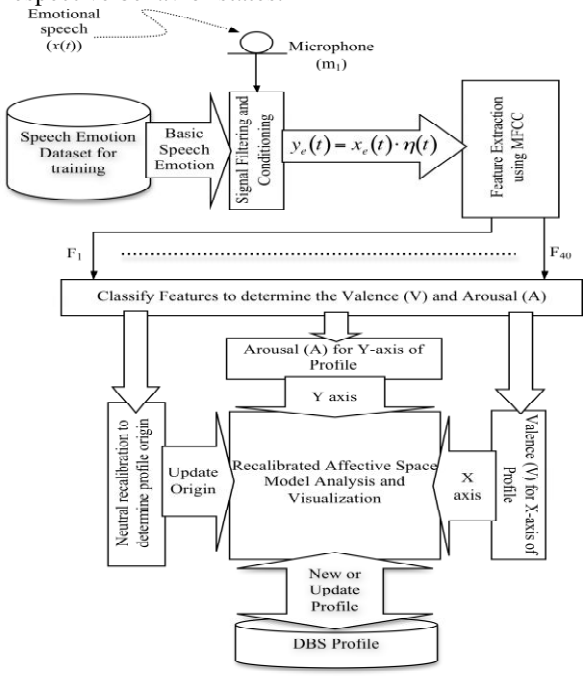

Figure 1. General Approach of the Proposed Driver Behavior State Profiling based on Recalibrated Affective Space Model.

Most psychologists agree that neutral is originated by the coordinate of (0,0) of the affective space model [6, 7]. However, such situation is almost impossible to achieve as other parameter, for instance culture, gives impact to the axis balance. Hence, we proposed that the centroid of neutral in the cultural-influenced speech emotion dataset to be used as the origin of the rSASM. The four quadrants of the rSASM are defined as hard boundary derived from the origin as a straight line separating between the positive and negative value of the valence and arousal (V,A) axes.

In order to find the emotion and neutral centroid, k-means clustering method was adopted. K-means clustering treats each instance as a point with its own coordinate. K-means finds the centroid by computing the distance between points and the cluster center iteratively so as to minimize the within-cluster sum of squares. Manhattan distance technique (L_1 distance) is employed. Each centroid is the component-wise median of the points in the neutral cluster.

Consider Manhattan distance, d_1 between 2 vectors of p and q in 2-dimensional affective space model with fixed Cartesian coordinate system. Centroid location can be calculated using Equation (1) where the result is the sum of the lengths of the projections of the line segment between the points onto the coordinate axes.

$$d_1(p,q) = \|p-q\|_1 = \sum_{i=1}^{2} |p_i - q_i| \quad (1)$$

where $p = (p_1, p_2, \ldots, p_n)$ and $q = (q_1, q_2, \ldots, q_n)$ are vectors.

Once the origin and the quadrant of the affective space model has been established, the boundary between emotion and neutral need to be determined. This neutral-emotion boundary refers to the soft boundaries segrerating neutral and emotion in its respective quadrant based on the assumption that the testing instant belongs to the class with nearest distance of either emotion centroid or neutral centroid in its respective emotion quadrant. The testing instance i distance to emotion centroid $d_{i,e}$ or neutral centroid $d_{i,N}$ can be computed by Equation (2).

$$d_{i,e} = \sqrt{(x_i - x_e)^2 + (y_i - y_e)^2}$$
$$d_{i,N} = \sqrt{(x_i - x_N)^2 + (y_i - y_N)^2} \quad (2)$$

where x_i and y_i are the valence and arousal values derived from the classifier outputs, x_N and y_N are the valence and arousal values of the neutral centroid and x_e and y_e are the valence and arousal values of the respective emotion centroid. Thus the instance i is more likely to be a member of a particular class if it has the nearest distance between the neutral and respective emotion centroid as illustrated by Equation (3), where the boundary between neutral and the respective emotion can be derived.

$$B = \frac{d_{i,e}}{d_{i,N}} = \begin{cases} >1 & \text{instance } i \text{ is neutral} \\ =1 & \text{this is the boundary and instance } i \text{ is taken as neutral} \\ <1 & \text{instance } i \text{ is the respective emotion} \end{cases} \quad (3)$$

To facilitate visualization, an ellipse based on scaling the Eigenvector to a unit circle is drawn representing the probability distribution of the scattered data for a particular set of HBS profiling on the rSASM. Detailed analysis of the HBS profile from the emotion primitives' based on both homogeneous and heterogeneous intra-cultural data arrangements are presented in Figures 2 to 4.

IV. Experimental Results and Discussion

Experimental result of the HBS profiling on the rSASM is presented in Fig. 2 and 3. It can be observed in these two figures that the normal HBS is densely populated at the center of the model. The distribution spread is more focused in Fig. 2 than Fig. 3 when we compare the normal HBS. This is because once we combined the three datasets as one big training data, the inter-cultural effect is suppressed.

Therefore, Normal HBS in Fig. 3 is almost evenly distributed compare to Fig. 2 that is biased towards Quadrant 1. It is also interesting to discern that sleepy HBS is largely populated in the negative region of arousal (y-axis). Such finding is inline with the psychologists understanding that sleepy HBS is a passive state. However, sleepy HBS seems to occupy Quadrant 3 indicating negative valence as opposed to the notion that sleepy falls in the Quadrant 4. The only logical explanation is that the individual starts to feel frustrated or irritated as he/she is forcing himself/herself to stay awake and alert.

The distracted and laughed HBS results are scattered mostly in the positive arousal region. This result indicates that the drivers are mostly active and engaged with the activities. Although distracted HBS is predicted to populate Quadrant 2 (drivers feel irritated when disturbed), some of the interruptions are welcomed resulting in positive valence (quadrant 1). The laughed HBS profiles are expected to be biased towards positive valance and are highly distributed in Quadrant 1 in both figures. Such result complements the psychologists' understanding that laughter is a cue to manifest happiness for speech-only situations.

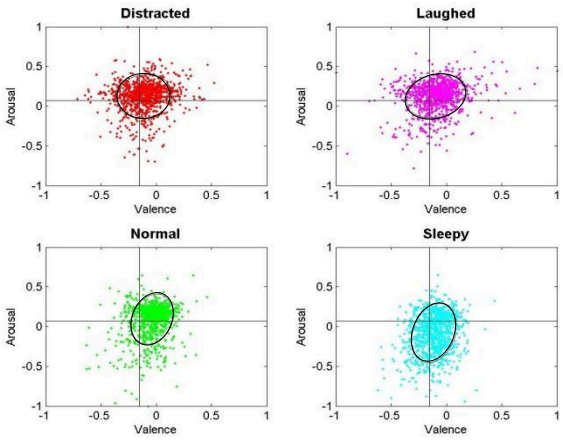

Figure 2. Human Behavior State Distribution using the Recalibrated Affective Space Model based on Homogeneous NTU_Asian Dataset.

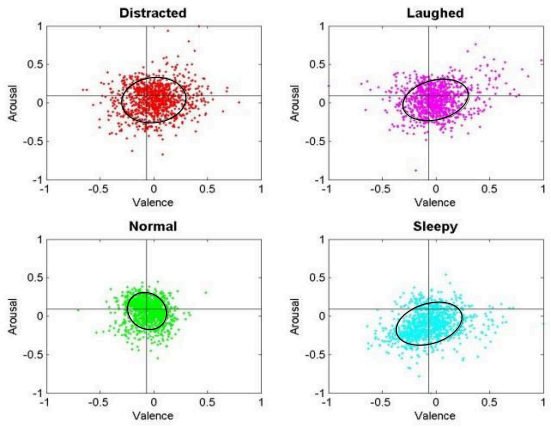

Figure 3. Human Nehavior State Distribution using the Recalibrated Affective Space Model based on Heterogeneous NTU_Asian + Berlin + NTU_American Datasets.

V. Conclusion

Based on the experimental results provided in Section 4, the HBS profiling based on rSASM enables researchers and medical practitioners to visualize the emotion primitives dynamically. From Figures 2 and 3 it is obvious that emotion dynamics are not discrete. Such approach can be used to show the emotion trajectory from one emotion to another in a more comprehensive manner.

More works are needed to fine-tune the results to improve the accuracy and usefulness in applying such approach in our daily life. An interviewer/ counselor could use a microphone with this approach in understanding and analyzing the subject's emotion primitives' dynamics to provide a more effective intervention.

References

[1] Z. Callejas and R. Lopez-Cozar, "Influence of Contextual Information in Emotion Annotation for Spoken Dialogue Systems", *Speech Communication*, Vol. 50, No. 5, 2008, pp. 416 – 433.

[2] T. L. Nwe, S. W. Foo and L. C. De Silva, "Speech Emotion Recognition Using Hidden Markov Models", *Speech Communication*, Vol. 41, No. 4, 2003, pp. 603-623.

[3] R. Cowie and R. R. Cornelius, "Describing the Emotional States That Are Expressed in Speech", *Speech Communication-Special Issue on Speech and Emotion*, Vol. 40, No. 1-2, 2003, pp. 5-32.

[4] R. R. Cornelius, "*The Science of Emotion: Research and Tradition in the Psychology of Emotion*", Upper Saddle River, NJ: Prentice-Hall. (1996).

[5] N. Kamaruddin, A. Wahab, and Q. Chai, "Cultural Dependency Analysis for Understanding Speech Emotion", *Journal of Expert System with Application (ESWA)* Vol. 39, No. 5, Apr 2012, pp.5115-5133

[6] P. J. Lang, "The Emotion Probe: Studies of Motivation and Attention", *American Psychologist*, Vol. 50, No. 5, 1995, pp. 372-385.

[7] H. Schlosberg, "Three Dimensions of Emotion", *Psychological Review*, Vol. 61, No. 2, 1954, pp. 81-88.

[8] F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier, and B. Weiss, "A Database of German Emotional Speech", *In: Proceeding of INTERSPEECH'05*, Lisbon, Portugal, 2005, pp. 1517-1520

[9] N. Kamaruddin and A. Wahab, "Heterogeneous Driver Behavior State Recognition using Speech Signal", *In: Proceeding of the 10th WSEAS International Conference on System Science and Simulation in Engineering (ICOSSSE '11)*, 3 – 5 October 2011, Penang, Malaysia, pp. 207-212.

[10] N. Kamaruddin, and A. Wahab, "Driver Behavior Analysis Through Speech Emotion Understanding", *In: Proceeding of the 2010 IEEE Intelligent Vehicle Symposium (IV 2010)*, 21-24 June 2010, San Diego, California, USA, pp. 238-243,

[11] J. A. Russell, "Affective Space is Bipolar", *Journal of Personality and Social Psychology*, Vol. 37, No. 3, 1979, pp. 345-356.

[12] M. Slaney, "Auditory Toolbox. (Ver. 2)", *Technical Report #1998-010*, Interval Research Corporation. (1998). [Online]. Available: http://cobweb.ecn.purdue.edu/~malcolm/interval/1998-010/

[13] N. Kamaruddin and A. Wahab, "Features Extraction for Speech Emotion", *Journal of Computational Methods in Science and Engineering (JCMSE)*. Vol 9, Supplement 1, (2009). pp. S1 – S12

[14] J. –S. R. Jang, "ANFIS: Adaptive-Network-Based Fuzzy Inference System", *IEEE Transaction on Systems, Man and Cybernetics*, Vol. 23, No. 3, 1993, pp. 665-685.

[15] M. Khalid, A. Wahab and N. Kamaruddin, "Real Time Driving Data Collection and Driver Verification Using CMAC-MFCC", *In: Proceeding of the 2008 International Conference on Artificial Intelligence (ICAI '08),* 14-17 Jul 2008, Las Vegas, Nevada, USA, pp. 219-224.