Development of EEG-based Emotion Classification Using Deep
Learning Networks

Mr. Suwicha Jirayucharoensak

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy (Computer Engineering) in
Computer Engineering
Department of Computer Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2018
Copyright of Chulalongkorn University

การพัฒนาระบบจำแนกอารมณ์จากสัญญาณคลื่นไฟฟ้าสมองด้วยโครงข่ายการเรียนรู้เชิงลึก

นายสุวิชา จิรายุเจริญศักดิ์

Thesis Title   Development of EEG-based Emotion
                            Classification Using Deep Learning Networks
By   Mr. Suwicha Jirayucharoensak
Field of Study   Computer Engineering
Thesis Advisor   Associate Professor Setha Pan-ngum, Ph.D.
Thesis Co Advisor   Pasin Israsena, Ph.D.

Accepted by the Faculty of Engineering, Chulalongkorn University in Partial Fulfillment of the Requirement for the Doctor of Philosophy (Computer Engineering)

.... Dean of the Faculty of Engineering
(Professor SUPOT TEACHAVORASINSKUN, Ph.D.)

DISSERTATION COMMITTEE
.... Chairman
(Professor Prabhas Chongstitvatana, Ph.D.)
.... Thesis Advisor
(Associate Professor Setha Pan-ngum, Ph.D.)
.... Thesis Co-Advisor
(Pasin Israsena, Ph.D.)
.... Examiner
(APIWAT LEK-UTHAI, Ph.D.)
.... Examiner
(Assistant Professor PEERAPON VATEEKUL, Ph.D.)
.... External Examiner
(Assistant Professor Yutana Jewajinda, Ph.D.)

สุวิชา จิรายุเจริญศักดิ์ : การพัฒนาระบบจำแนกอารมณ์จากสัญญาณคลื่นไฟฟ้าสมอง
ด้วยโครงข่ายการเรียนรู้เชิงลึก. (Development of EEG-based
Emotion Classification Using Deep Learning Networks)
อ.ที่ปรึกษาหลัก : รศ. ดร.เศรษฐา ปานงาม, อ.ที่ปรึกษาร่วม : ดร.พศิน อิศรเสนา ณ
อยุธยา

กระบวนการจำแนกอารมณ์เป็นหนึ่งในองค์ประกอบสำคัญของการปฏิสัมพันธ์ระหว่างเครื่องคอมพิวเตอร์กับ
มนุษย์ (Human Computer Interaction) ทำให้เครื่องคอมพิวเตอร์ปฏิสัมพันธ์กับมนุษย์ได้อย่างมีประสิทธิภาพมาก
ขึ้น สัญญาณคลื่นไฟฟ้าสมอง (EEG) เกิดจากการทำงานร่วมกันของเซลล์ประสาท (Neural Cells) เป็นลักษณะขั้นลำดับ
ที่ซับซ้อน ที่มีความเกี่ยวข้องโดยตรงกับสภาวะทางอารมณ์ ดังนั้นการจำแนกอารมณ์จากสัญญาณคลื่นไฟฟ้าสมองต้องอาศัย
กระบวนการเรียนรู้ที่มีความสามารถในการสร้างโมเดลระดับสูง (High-level Abstraction) ได้ งานวิจัยชิ้นนี้ได้
ประยุกต์ใช้โครงข่ายการเรียนรู้เชิงลึก (Deep Learning Network) เพื่อช่วยเพิ่มประสิทธิภาพทางด้านความแม่นยำของ
ระบบจำแนกอารมณ์จากสัญญาณคลื่นไฟฟ้าสมอง โดยอาศัยความสามารถในการเรียนรู้ลักษณะลำดับขั้น (Hierarchical
Learning) ซึ่งเหมาะสมกับการเรียนรู้ข้อมูลจากสัญญาณคลื่นไฟฟ้าสมอง ต่อมางานวิจัยชิ้นนี้ได้ศึกษาองค์ความรู้ทางด้านการ
เปลี่ยนแปลงการทำงานของสมองในสภาวะทางอารมณ์ต่างๆ (Temporal Neural Dynamics of Emotions)
และได้พัฒนาต่อยอดระบบให้มีความสามารถในการเรียนรู้การเปลี่ยนแปลงที่สอดคล้องกับสภาวะทางอารมณ์ต่างๆ โดยอาศัย
การทำงานของ Hidden Markov Model (HMM) จากผลการทดสอบพบว่า การทำงานร่วมกันของ DLN กับ
HMM มีประสิทธิภาพในการจำแนกอารมณ์จากสัญญาณคลื่นไฟฟ้าสมองดีขึ้น เมื่อเปรียบเทียบกับการทำงานของ DLN
อย่างเดียว โดยค่าความแม่นยำในการจำแนก Valence 3 ระดับ เพิ่มจาก 60.07% เป็น 64.18% และค่าความแม่นยำใน
การจำแนก Arousal 3 ระดับ เพิ่มจาก 59.83% เป็น 62.98%

สาขาวิชา   วิศวกรรมคอมพิวเตอร์   ลายมือชื่อนิสิต ....
ปีการศึกษา   2561   ลายมือชื่อ อ.ที่ปรึกษาหลัก ....
                                                      ลายมือชื่อ อ.ที่ปรึกษาร่วม ....

5671431421 : MAJOR COMPUTER ENGINEERING

KEYWOR
D:
Emotion Classification, Human Computer Interaction, Deep Learning Networks, Electroencephalogram

Suwicha Jirayucharoensak : Development of EEG-based Emotion Classification Using Deep Learning Networks. Advisor: Assoc. Prof. Setha Pan-ngum, Ph.D. Co-advisor: Pasin Israsena, Ph.D.

Emotion classification is one of essential tasks for Human Computer Interaction (HCI) to make computers more efficiently interact with their users. Electroencephalogram (EEG) signals, associated with cognitive states of emotions, propagate though a complex hierarchy of neuron cells. Therefore, EEG-based emotion classification requires sophisticated learning algorithms that can represent high-level abstraction of a complicated task. This dissertation focuses on applying deep learning networks (DLNs) to enhance accuracy performance of the EEG-based emotion classification system. The DLN provides hierarchical feature learning methodology which is suitable for EEG-related feature learning algorithms. Furthermore, this research investigates the effects of temporal neural dynamics of emotions and then focuses on learning state transitions of DLN's high-level features by using hidden markov models (HMMs). From experimental results, our proposed EEG-based emotion classification system with hybrid DLN-HMM has better accuracy performance compared with using only DLNs. The average of classification accuracy for 3-class valence improves from 60.07% to 64.18% and those of 3-class arousal improves from 59.83% to 62.98%.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Field of Study:	Computer Engineering	Student's Signature	....
Academic Year:	2018	Advisor's Signature	....
		Co-advisor's Signature	....

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

จุฬาลงกรณ์มหาวิทยาลัย

CHULALONGKORN UNIVERSITY

LIST OF FIGURES

Chapter 1

Introduction

1.1 Motivations

Nowadays brain computer interface (BCI) is one of the breakthrough research fields in biomedical engineering. Many inexpensive EEG amplifier headsets are available in the market for non-invasive EEG signal acquisition. Most of EEG-based BCI algorithms have been designed for controlling devices by regulating user's brain wave activities. There are many of competent BCI-based applications such as word typing programs [1] and wheelchair controllers [2]. Not only can BCI technology be used to manipulate computers or devices, but it can be implemented for estimating our cognitive states of emotions as well.

Emotion state recognition is a challenging application of Human Computer Interaction (HCI). To make computers or machines more efficiently interact with their users, they are required to recognize emotion states of the users more precisely. EEG-based emotion classification systems require sophisticated learning algorithms that can represent high-level abstraction. During the decade, the idea of deep learning algorithms has been intensively applied to a variety of complicated tasks such as vision recognition [3, 4] and natural language processing [5, 6]. The results from the DLNs outperform the previous state-of-the-art techniques of these applications. The original concept of greedy layer-wise unsupervised pre-training in the deep learning algorithm was invented by Hinton et al. [7] and its deep network was implemented with a stack of restricted boltzmann machines (RBMs) . They demonstrated the effectiveness of unsupervised pre-training technique by performing handwritten digit recognition in MNIST database. Subsequently, Bengio et al. [8] applied the unsupervised pre-training algorithm to a stack of autoencoders. This work provided a new mechanism to naturally handle continuous-valued inputs providing better predictive models.

The DLN concept can be applied to EEG-related applications. Wulsin et al. [9] applied the unsupervised pre-training concept to a stack of autoencoders for classifying and indicating abnormal patterns in EEG signals. The aim of this study is to develop a classifier used to perform specific waveform rate estimations of EEG signals from each electrode channel. The authors compared the performance of four different classifiers: Decision Tree (with Gini diversity index), SVM (with a radial basis kernel function), K-NN (with Euclidean distance) and DLN (with 4-layer RBMs), by using raw data and extracted features. Experimental results showed that DLNs provided more efficient for recognizing anomaly EEG waveforms than other methods.

Moreover, the DLN can be applied to perform sleep stage classification [10]. This work utilized both raw EEG signals and power spectral features for an unsupervised feature learning. The study showed how to apply the concept of unsupervised feature learning to eliminate the use of handcrafted feature selection. In this study, there are three experimental setups: DBN with raw data, DBN with feature extraction and DBN with PCA-feature selection. Experimental results revealed that the deep architecture using raw data provided comparable performance to the feature-based approach.

The deep learning network (DLN) is responsible for discovering unknown feature coherences between input signals that is significant to represent a complicated model [11]. More generally, the DLN potentially eliminates the limitation of handcrafted feature selection, which is considerably difficult to determine. Moreover, the DLN provides hierarchical feature learning approach in which high-level features can be learnt from compositions of low-level features with greedy layer-wise unsupervised pre-training. Consequently, the DLN may be useful to enhance feature learning and classification algorithms for emotion state estimation because it performs hierarchical feature learning which is suitable for EEG-related algorithms. The brain waves or EEG activities, which are associated with cognitive functions, propagate through a complex hierarchy of neuron cells [12].

1.2 Objectives

The main purpose of this dissertation is to study the feasibility of utilizing DLNs to perform hierarchical feature learning algorithms for EEG-based emotion classification. This dissertation further examines the temporal neural dynamics of emotion processes and then develops a hybrid DLN-HMM system to enhance its accuracy performance by learning relevant state transitions of DLN's output features with HMMs.

1.3 Contributions

There are several principle contributions of this dissertation as following.

- Explore the feasibility of Deep Learning Networks to implement EEG-based emotion classification systems and then apply the concept of principal component-based covariate shift adaptation to alleviate non-stationary effects of EEG signals.

- Introduce a hybrid DLN-HMM approach to learn the effects of temporal neural dynamics of emotion processes and evaluate the impacts of key configuration parameters of our proposed hybrid system.

- Examine the temporal dynamics of EEG power correlations with emotion states.

1.4 Chapter Organization

The rest of this dissertation is organized as following chapters:

Chapter 2: This chapter provides related theories applied to develop our proposed systems. This chapter starts with methods of emotion classification using physical and physiological signals. Next, EEG correlates with emotions are described and then following with overview of EEG-based emotion classification. Subsequently, we explain the fundamental theory of DLN in order to provide some background knowledge of hierarchical layer-wise learning and training protocols. Finally, Hidden Markov Model (HMM) and Temporal neural dynamics of emotions are described how they can be applied to learn state transitions of our proposed system.

Chapter 3: Literature reviews of relevant researches for developing EEG-based emotion classification systems are provided. This chapter also describes several researches using EEG and other physiological signals to enhance accuracy performance of emotion state classification. Next, relevant research studies of EEG-based emotion classification using DLNs were discussed in order to show how the DLN could be applied for EEG-based emotion classification systems.

Chapter 4: This chapter presents our first approach to develop an EEG-based emotion classification system with DLNs. The system is implemented with a stack of autoencoders and softmax output classifiers. In addition, the concept of covariate shift adaptation of principle components is applied to alleviate non-stationary effects of EEG signals. In this approach, all labels of training and testing data samples deriving from one emotion stimulus are marked as the same emotion class. Therefore, we simply call this approach as "Static DLN".

Chapter 5: This chapter introduces a hybrid DLN-HMM model to develop a novel EEG-based emotion classification system. We exploit the concept of temporal neural dynamics of emotions to enhance its accuracy performance by implementing HMMs to learn all relevant state transitions of high-level DLN features. The HMMs use the posterior probabilities of state sequences to classify the cognitive states of emotions. In this approach, the labels of training and testing data samples are marked as a sequence of targeted hidden states deriving from GMM-HMM models. We call this approach as "Dynamic DLN-HMM".

Chapter 6: The purpose of this chapter is to compare the accuracy performance of two approaches – static DLN vs dynamic DLN-HMM. We perform statistical analysis by computing paired t-test of their classification accuracy values with network parameter variants -- Number of hidden DLN layers and Number of input frame window.

Chapter 7: This chapter examines EEG power correlations with emotions and reveals relevant spectral power synchronizations/desynchronizations at particular electrode locations. Moreover, we perform time-course analysis of spectral powers to discover some common patterns of EEG power correlations with emotions for individual subjects.

Chapter 8: The last chapter provides conclusions of this research study. All key knowledge discovered from experiments in static DLN and dynamic DLN-HMM approach are summarized. Furthermore, potential future works of this research study are discussed.

Chapter 2
Related Theories

2.1 Methods of Emotion Classification

Emotion classification is one of the most challenging algorithms for Human Computer Interaction (HCI) applications. To develop machines or computers to be more efficiently interact with human beings, they must be able to recognize user's cognitive states of emotions more precisely. The tasks will bridge the gap between computers and human beings for considerable improvement of interactions. Emotion classification or recognition can be categorized into two main methods by using physical and physiological signals.

The first method to recognize emotions is to employ physical signals such as facial expressions, body gesture and speeches. There are a plenty of researches investigating how to perform emotion classification by using these physical signals. Emotions can be recognized by facial expression. Facial emotions are crucial characteristics that help us to recognize the sensations of others. We are able to sense the emotion states of other people using facial expressions. In general, facial emotion recognition (FER) are consisted of three steps – (1) facial component detection, (2) feature extraction and (3) expression classification, illustrated in Figure 1 [13]. In the first step, the system detects faces appeared on input images and then extracts a variety of both spatial and temporal features from the facial components. Next, the pre-trained FER algorithms, such as a Support Vector Machine (SVM), AdaBoost and random forest, predict the state of emotions. Walecki et al [14] implemented a deep-learning based FER system by using convolutional models (CNN) without handcrafted features. The CNN-based FER system is able to learn all features hierarchically to produce feature maps. Then, each feature map is fed to a series of fully connected networks and then softmax layers estimate the facial expressions.
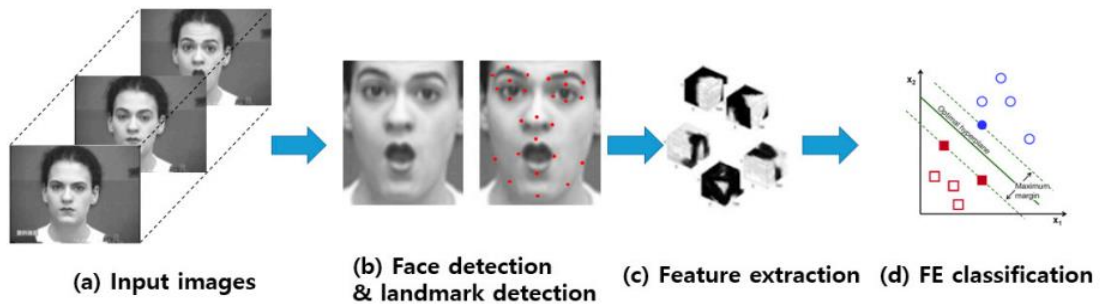
Figure 1   Procedure in FER system.

One of affective expression recognitions employs body gesture forms and movements. Kleinsmith et al [15] provides an extensive literature review of affective body expression perception and recognition. A number of researches studied the relationship between body gesture and affective expression. Most of these researches examined the correlations of body skeleton with each emotion by using a variety of mapping algorithms. Kapur et al [16] used full body skeleton movements captured by video-based sensors to perform training on five learning algorithms – logistic regression, naïve bayes, decision tree classifier, multi-layer neural network and support vector machine. The features include velocity and acceleration of relevant masks in the skeleton. Glowinski et al [17] investigated movement of upper body (head and hands) from video captured from ten professional actors. The movement (velocity and acceleration) of a head and hands was calculated on the basis of the coordinates of the blob's centroids using a proprietary program. Ravindra et al [18] is to investigate the possibility of grounding nuances of affective states on human postures. The study applied Mixed Discriminant Analysis (MDA) to the numerical description of the collected postures for mapping the postures on to a multidimensional discriminant space. Figure 2 [18] shows 3D affectively expressive avatars for each nuance of emotion (a) Upset (b) Angry (c) Happy (d) Joyful (e) Low-intensity Surprise (f) Fear (g) High-intensity Surprise (h) Sad and (i) Depressed.
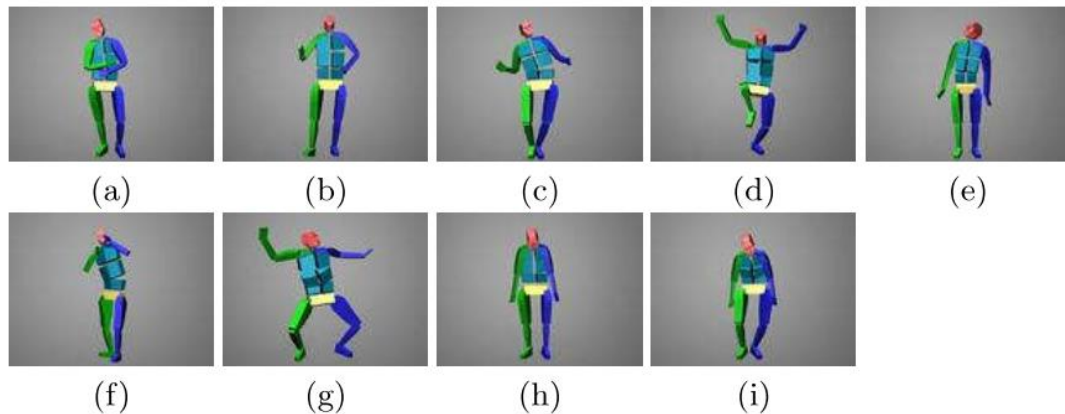
Figure 2   Affectively expressive avatars

The cognitive states of emotions can be recognized from human's voice or speech. Human's speech is a complex physical signal containing information about language. Not only speech delivers the information of messages but it simultaneously sends the sensation of messages as well. Emotion classification or recognition from speech is a challenging task. Most existing speech-based emotion classification provide poor performance because the difficulty in modeling and characterization of emotions presented in speech. Hosseini et al [19] employed Gaussian Mixture Model (GMM) to learn MFCC features of the speech. In training phase, the system uses Kullback-Leibler Divergence (KLD) as a measure to detect the discriminative GMM mixtures. Each GMM model represents one of the cognitive states of emotions. In testing phase, the discriminative frames are recognized based on Frame Selection Decoding (FSD), shown in Figure 3 [19]. Li et al. [20] developed a speech emotion classification in hybrid DNN-HMM approach. The method uses Deep Neural Network (DNN) to hierarchical learn MFCC features derived from GMM-HMM models. Each GMM-HMM represents the model of each emotion. The result of emotion classification will be the model that provides the maximum likelihood of GMM-HMM model.
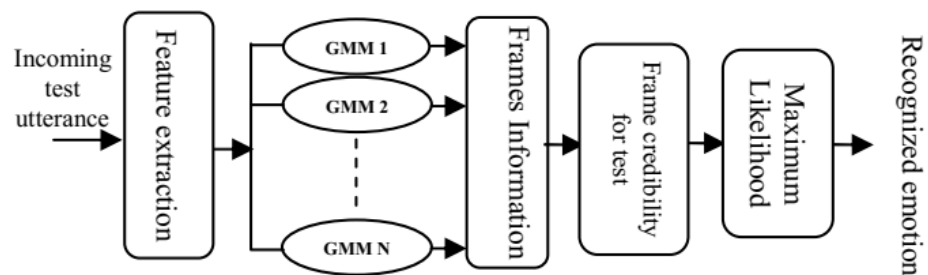
Figure 3   GMM-based Emotion Classification from Speech

The second approach to recognize emotions is to use physiological signals such as electrocardiogram (ECG or EKG), respiration (RSP) and galvanic skin response (GSR). Emotions are complex processes comprised of numerous components, including physiological signal changes. This emotion recognition approach uses bio-signals to provide more accurate emotion estimation. Subjects can pretend their physical signals (facial expression and speech) but they cannot pretend their physiological signals. The transition of the emotion states affects Electrocardiogram or ECG signals. Therefore, ECG signals are potentially effective indicator to recognize emotions.

There are relevant research studies on ECG-based emotion classification. Valenza et al.[21] proposed an innovative personalized probabilistic framework to characterize the emotion states through the analysis of heartbeat dynamics exclusively. The system performs artifact removal and band-pass filtering. Then, RR interval series are extracted by using automatic R-peak detection algorithms. Next, both linear and non-linear features are extracted by using Nonlinear Autoregressive Integrative (NARI) model. A SVM classifier achieves an overall accuracy for recognizing four emotional states 79.29%, with 79.15% on the valence axis, and 83.55% on the arousal axis. Cheng et al [22] presents a novel method on negative valence detection via feature fusion form one-channel real-time ECG signal. A set of features was extracted from raw ECG signals and heart rate variability (HRV), including linear features, non-linear features, time-domain features and time-frequency domain features, shown in Figure 4 [22]. All of these features are fed into a SVM classifier with 79.51% accuracy.
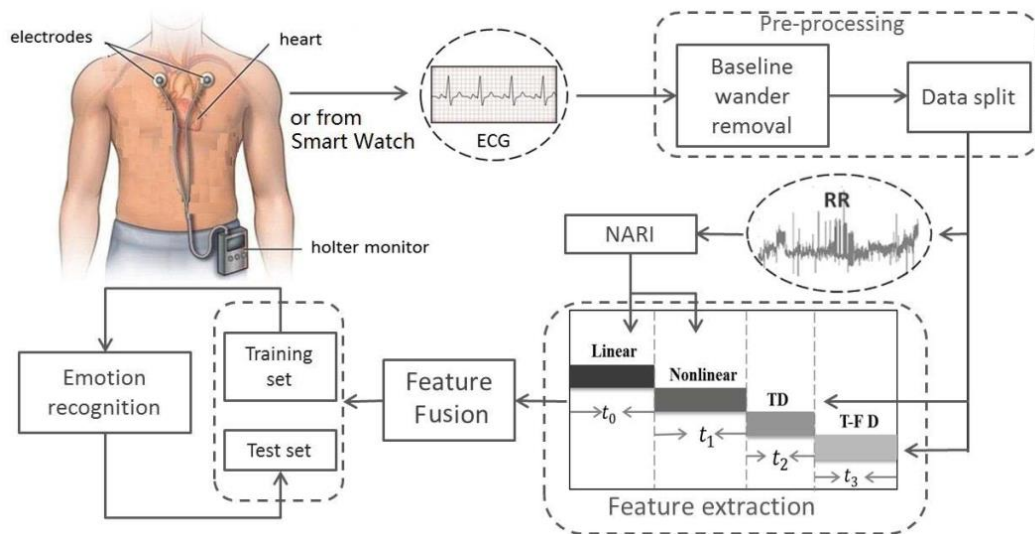
Figure 4   ECG-based Emotion Classification

One of physiological signals using for emotion recognition is respiration (RSP) signals. The RSP signals are relevant to the cognitive states of emotions. Wu et al [23] presents an automatic respiration signal segmentation method to extract representative Emotion Elicited Segments (EESs) for emotion states of individual subjects. The EES process consists of Mutual Information-Based Emotion Relevance Feature Ranking and Constraint-based Elicited Segment Density (CESD) analysis. Experimental results obtained for five prototypical emotions show that the proposed segmentation/ extraction methodology provides an average classification rate of 88%. Mirmohamadsadeghi et al. [24] used the combination of RSP and ECG signals to enhance the efficiency of emotion classification. Features are a set of separate RSP and ECG features as well as synchronization aspects of both signals. The algorithm has three different binary classifications (liking, arousal and valence) and an SVM classifier was used for the classification of a single trial. Then a majority voting on these single-trial classification results was performed as depicted in Figure 5 [24]. Experimental results show the classification accuracy is approximately 71%.
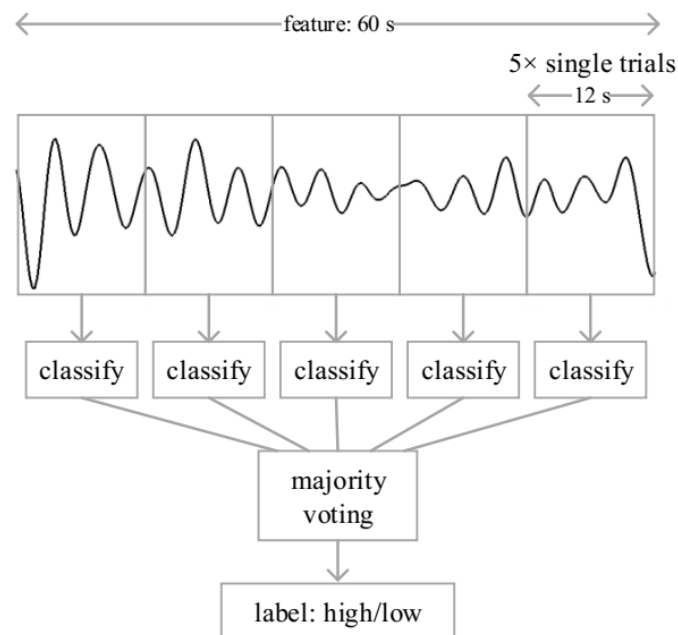
Figure 5   *Classification procedure for a single trial*

One of the most sensitive measures for emotion states is Galvanic Skin Response (GSR). The GSR changes from sweat glands in our skin and reflects the intensity of emotion. Liu et al, in Figure 6 [25], proposed a GSR-based Emotion classification with covariance-based feature selection. The proposed algorithm selected 15 out of 30 original GSR features. A SVM classifier performed 5-class emotion (happiness, grief, fear, anger, and calm) recognition with 66.67% accuracy. A research [26] used the combination of GSR and ECG signals to perform emotion classification with Matching Pursuit (MP) algorithm. These two signals collected from 11 healthy students while subjects were listening to emotional music clips. The proposed algorithm applied three dimensionality reduction methods, including Principal Component Analysis (PCA), Linear Discriminant Analysis, and Kernel PCA. These size-reduced features were inputs of a PNN classifier. The PCA approach provided the maximum accuracy of 92.52%.
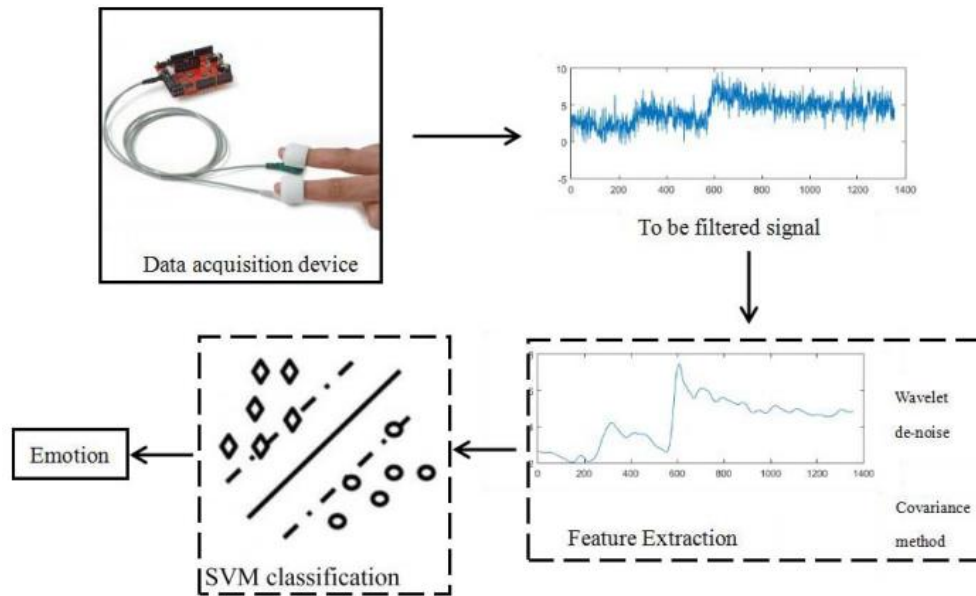
Figure 6   GSR-based Emotion Classification

2.2 EEG Correlates of Emotion

Emotion involves the entire of nervous system of our body, especially brain. The limbic system is primarily responsible for generating and regulating our emotional life. Another function of the limbic system is to help us form and retain our memories, which is crucial for learning and developing process of human's life. The limbic system is a complicated brain structures that locates under the cerebrum and beside of the thalamus.

The limbic system is constituted of three main components: Hypothalamus, Amygdala and Hippocampus, illustrated in Figure 7 [27].
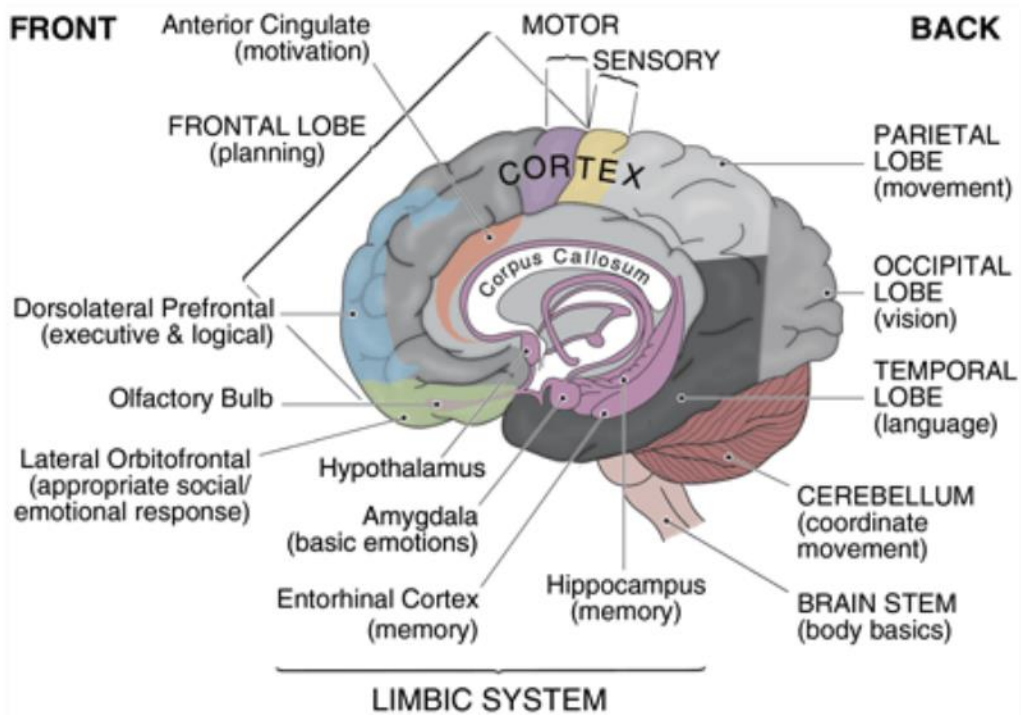
Figure 7   Limbic System of Brain Structure

Hypothalamus is responsible for processing and regulating emotion. It plays an essential role in sensory processes by receiving all sensory inputs such as vagus nerve, reticular formation, optic nerve and olfactory nerves. Then, it commands a set of instructions to control our body through Autonomic Nervous System (ANS). It allows the hypothalamus to regulate our sympathetic and parasympathetic functions.

Amygdala is mainly capable of controlling our instantaneous emotional responses. It also manipulates our judgments, discerning between good and bad. With memory association, it helps the brain recognize potential threats and prepare for protection. The amygdala also performs learning processes based on reward or punishment criteria.

Hippocampus performs an essential role for new memory formation and organization. It takes in charge of collaboration of emotion and memory by connecting certain sensations to the memory cells in this area. These connections help us recall associated memories with various states of cognition. Moreover, the hippocampus is crucial for our spatial orientation and helps us to recognize navigation directions. In addition, there is another part of the brain involving emotion experience. The prefrontal cortex, located at frontal lobe region, collaborates with the limbic system to regulate our emotional life. It is mainly responsible for thinking, making plans, and taking action. However, with modulation of Amygdala reactivity, the prefrontal cortex plays an essential role to control our emotion and behave with appropriate manners [27].

2.3 EEG-based Emotion Classification

Electroencephalogram (EEG) is the oscillation of electric potentials occurring from miniature amount of current flows between synapses in brain. EEG signals can be obtained by measuring the electric potentials at positions that electrodes installed on the scalp. The 10-20 international system of electrode placement, illustrated in Figure 8 [28], provides a common system to perform measurement reproducibility. With 10-20 international system, a subject's studies can be comparable in time-course manners and between subjects. The brain wave is a combination of five bands -- Delta (1-4 Hz), Theta (4-8 Hz), Alpha (8-12 Hz), Beta (12-30 Hz) and Gamma (30-50 Hz), illustrated in Figure 9 [29]. The process of emotional state estimation intensively investigates the characteristics of each band for more reliable recognitions.
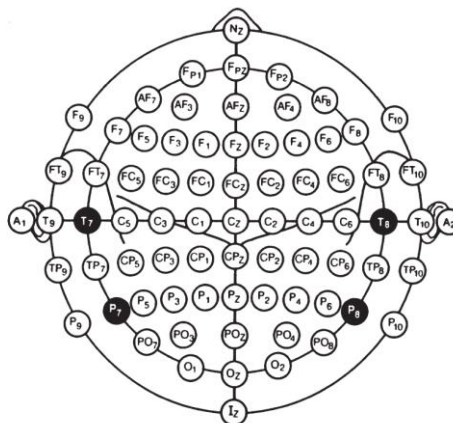
Figure 8   10-20 International System
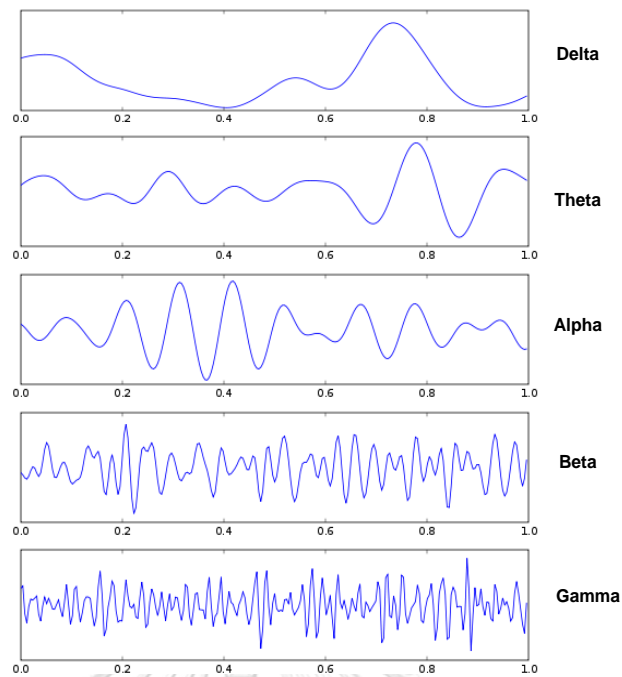
Figure 9   Brain waves in each frequency band

There are two fundamental methods to provide a definition of each emotion state. First, category-based emotion modeling approach [30-33] defines each state of basic emotions corresponding to facial expressions or bodily involvement such as Happy, Sad, Fear, Angry, etc. Another approach is multi-dimensional emotion modeling [34]. The cognitive states of emotions can be determined by two parameters -- valence and arousal. The valence is the level of attractive feelings that a person feels toward a particular object or event. The arousal is a state of awakeness to a particular stimulus. The two dimensional model of emotion states, depicted in Figure 10 [35], is more efficient to represent a variety of emotion states than category-based approach does. In this research study, we classify the cognitive states of emotions according to valence and arousal status from EEG signals.
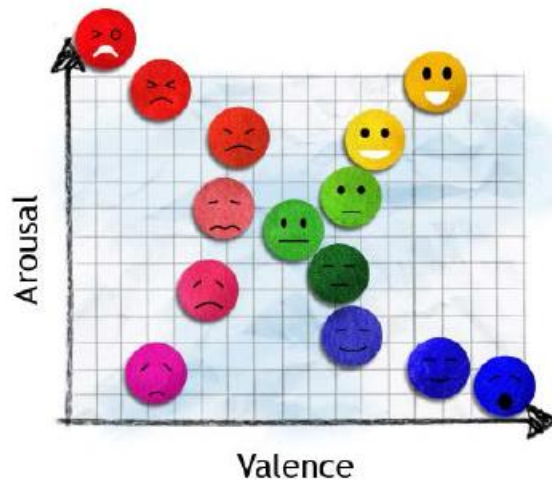
Figure 10   Valence-arousal dimensional model

2.4 Deep Learning Network

Deep Learning Network (DLN) is an evolution of Artificial Neural Network (ANN). Deep learning algorithm [36] exploits the concepts of hierarchical feature learning and unsupervised pre-training techniques to enhance its effectiveness of complex learning algorithms. Therefore, with the deep learning algorithm, we can implement ANNs with more layers for learning more complicated tasks. The DLN has potential to discover unknown feature coherences of inputs that is crucial to learn such a complex model. The key capability of the DLN is to perform hierarchy feature learning with unsupervised pre-training. With this approach, the features at its high-level can be investigated from compositions of features at its low-level with greedy layer-wise unsupervised pre-training. After this unsupervised pre-training completes, the network will be fine-tuned with gradient-based optimization. Therefore, the DLN is capable of learing complex tasks to represent high-level abstraction. A hierarchical architecture of the DLN is shown in Figure 11.
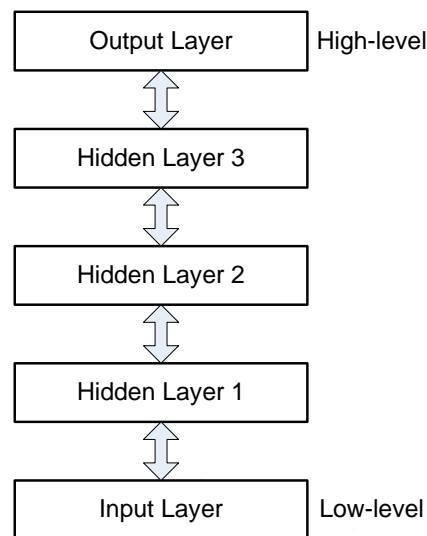
Figure 11   Hierarchical architecture of the DLN

Furthermore, the DLN has potential to perform self-taught learning from a tremendous number of unlabeled data. Usually, when learning algorithms investigate more information, they provide better performance. The prominent advantage of self-taught learning is that the algorithm can learn from unlabeled data. In other word, it can learn from a large amount of information. Consequently, the DLN is suitable for problems where there are plenty of unlabeled data and only a little amount of labeled data.

DLN training procedure is depicted in Figure 12. The procedure begins with a greedy layer-wise unsupervised pre-training process from the first hidden layer to the last hidden layer. Initial values of weights and biases are required for parameter optimizations at the current trained hidden layer. Subsequently, the features deriving from feedforward propagation must be used to perform unsupervised pre-training in the next hidden layer. After the last hidden layer completes unsupervised pre-training, softmax training and fine-tuning procedures need to be performed respectively.
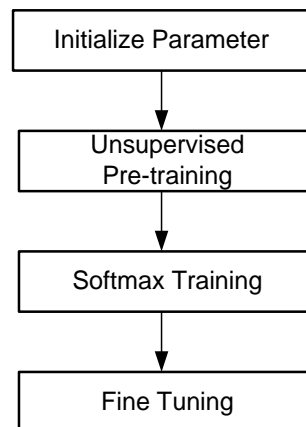
Figure 12   DLN training procedure

2.5 Markov Chain and Hidden Markov Model

Markov chain is a stochastic model representing discrete-state sequences of events in a state space ($q_t \in \{ s(j), \ j = 1, 2, \ldots, N\}$). With Markov assumption, the probability of current state only depends on its previous state. Markov chain, $q_1^T = q_1$, $q_2, \ldots, q_T$, is characterized by the probabilities in its transition matrix, defined by

$$P(q_t = s^{(j)} \mid q_{t-1} = s^{(i)}) = a_{ij}(t)$$

The Markov chain is considered as a time-sequence model to provide observation sequences with one-to-one relation to a state. Hidden Markov Model (HMM) is an extenstion from the Markov chain involving uncertainty by associating an observation probability distribution with each state in the Markov chain. Therefore, a HMM is a doubly stochastic process of the Markov chain. HMM state sequences can be only observed through the observation probability distribution, as shown in Figure 13 [38].
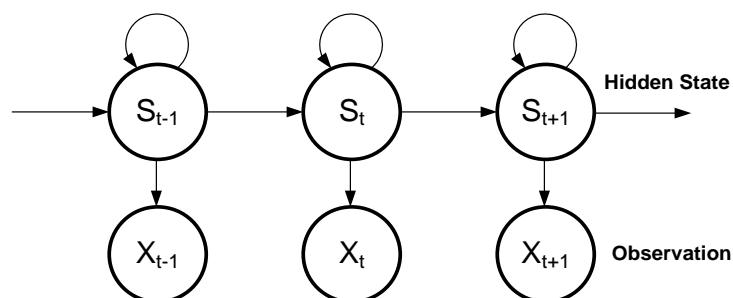
Figure 13   Hidden Markov Model

There are three configuration parameters of HMMs: Transition probability matrix (A), Initial state probabilities (¶) and Emission probability (B).

- Transition probability matrix, $A = (a_{ij})$; $i,j = 1, 2, ..., N$

$$a_{ij} = P(q_t = j \mid q_{t-1} = i)$$

- Initial state probabilities, $\P = [(\P_i)$; $i = 1, 2, ..., N$

$$\P_i = P(q_1 = i)$$

- Emission probability, $B = (b_i)$; $i = 1, 2, ..., N$

$$b_i(x_n) = P(x_n \mid q_n = s_i)$$

The process of learning HMM parameters employs the Baum-Welch algorithm, one of the efficient Expectation-Maximization (EM) algorithms, for estimating hidden variables in a statistical model. After completing the process of learning configuration parameters of HMMs, Viterbi algorithm can be applied for decoding optimal HMM state sequence associated with the given observation sequence.

2.6 Temporal Neural Dynamics of Emotions

Fundamental characteristics of emotions are congruous with multiple brain wave activities and interactions among them over time. Therefore, the concept of temporal neural dynamics of emotions is crucial to elucidate the brain representation underlying the perception of emotions. Emotional responses are derived from multiple interacting component processes such as appraisal, expression, conscious experience, physiological arousal and action tendencies [39].

The procedure of emotional activation in brain involves with three temporal dynamic features of an emotional response to a single event – Emotion onset, Emotion duration and Emotion Resurgence, illustrated in Figure 14 [40].
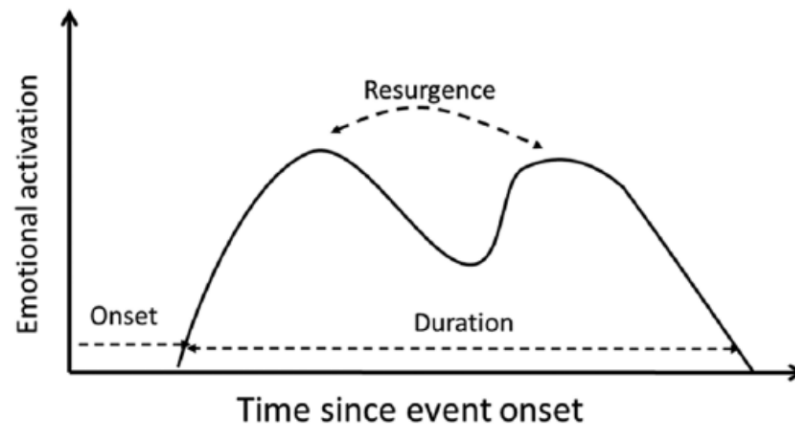
Figure 14   Temporal dynamic Features of an emotional response

The onset of neural responses to an emotional event or stimulus is the early integrative process of emotional appraisals, ranging from initial coarse evaluation to more complex evaluation of a stimulus. Most of the initial appraisals are automatic and spontaneous processes. The timing of initial appraisals for novelty occurs approximately 95-200ms after the stimulus, and before appraisals for pleasantness 200-300 ms. Finally, the onset process is completed with appraisals of stimulus relevance approximately 400ms.

The duration of emotional experiences is variable ranges of time, from a few seconds to several hours, after their initial onset. The key factor of emotion duration is the initial intensity of that emotion experience. A research [41] examined the neural correlates of working memory for emotional information. From an fMRI study, they found that participants had longer than 7 second for maintaining their emotion responses. A study [42] investigated the amount of time for emotion suppression. The study found that the emotion suppression, changing one emotion to another, requires 10.5 to 15 second.

The resurgence of emotion is a common situation that emotion activation has multiple peaks after a single event of emotion stimulus. There are several reasons to explain this emotion resurgence. First, the following peak may be a result of recollecting or rethinking about the emotion stimulus or event. Second, the multiple peaks of emotion activation may be derived from the combination of two different processes that occur in different times such as threat anticipation and threat imminence.

Chapter 3
Literature Reviews

There was a variety of research studies [43-49] to examine the efficiency of their implementation for EEG-based emotion classification systems with SVM classifiers. Summary of these implementation techniques is shown in Table 1. This research study desired to evaluate a benchmark of these implementations and found that it was difficult to compare their classification accuracy because these implementions employed a variety of cross validation approaches and feature selection techniques as well as different styles of emotion stimulus protocol. Most of these EEG-based emotion classification systems employed the SVM algorithm as their emotion classifiers. As we know, the limitation of the SVM classifiers is the curse of dimensionality, especially in real-time classification. These systems are required to perform feature selection in order to reduce the number of learned features prior to SVM classification.

Moreover, there exists several intensive research studies to enhance their emotion classification systems by using EEG signals and other physiological signals. Koelstra et al [50] proposed a method utilizing EEG and physiological signals. This work employed power spectral as its features and used a SVM classifier to classify 2-state of valence and arousal emotion. The proposed system provided the classification accuracy of 55.7% for arousal and 58.8% for valence. A work [51] presented a multi-modality dataset of emotion states. The combination of EEG and other physiological signals was used to classify the cognitive states of emotions. The proposed system employed EEG power spectral from 32-channel as input features. A SVM classifier was responsible for classifying 3-state of valence and 3-state of arousal emotions. For EEG-based classification, the accuracy performance for valence and arousal are 57.0% and 52.4% respectively. Dong et al.[52] developed an Asymmetry Spatial Pattern (ASP) technique to extract features. This work compared the accuracy performance of naive bayes, K-NN, and SVM algorithms for affective state classifications. The average accuracy for valence state is 66.05% and arousal state is 82.46%.

Table 1 Benchmark of EEG-based Emotion Classification Systems

Ref	#Subj	Emotion Classes	Stimulus	Classifier	Accuracy
[43]	10	2 Valence classes	Picture, Music	SVM	75.62%
[44]	32	2 Valence, 2 Arousal	Video	SVM	80.43% (Valence) 71.11% (Arousal)
[45]	26	4 Classes (joy, anger, sad, pleasure)	Music	SVM	82.29%
[46]	26	2 Valence, 2 Arousal	Music	SVM	76% (Valence) 74% (Arousal)
[47]	14	2 Valence	Sound	SVM	87.02%
[48]	6	2 Valence classes	Movie Clip	SVM	87.53%
[49]	5	4 Classes (joy, relax, sad, fear)	Movie	SVM	66.51%
[50]	6	2 Valence, 2 Arousal	Music Video	SVM	55.7% (Valence) 58.8% (Arousal)
[51]	27	3 Valence, 3 Arousal	Video	SVM	57.0% (Valence) 52.4% (Arousal)
[52]	4	2 Valence, 2 Arousal	Video	K-NN,NB, SVM	66.05% (Valence) 82.46% (Arousal)
[53]	7	2 Valence, 2 Arousal	Picture	SVM	73.42% (Valence) 73.57% (Arousal)
[54]	6	2 Valence	Video	DNN,GELM, SVM,K-NN	87.62%
[55]	21	4 emotion classes	-	DNN	68.4%
[56]	14	3 Valence	Video	DNN	91.36%

Most of EEG-based BCI applications confront with high level of inter-subject variations and non-stationarity of EEG signals. Relevant features for emotion classification are the characteristics of an individual subject and possibly change over time during experiments. Consequently, the EEG-based emotion classification requires a sophisticated feature learning technique such as deep learning algorithm to handle these limitations. The deep architecture can be employed to classify the cognitive state of emotions from EEG signals using DNNs [54]. The deep network is implemented with a stack of RBMs and its features were extracted with differential entropy (DE) for each frequency band. This work performs 2-class valence and compares the performance of the deep network to K-NN, SVM and GELM (Graph regularized Extreme Learning Machine). The average accuracy of DBN, GELM, SVM and K-NN are 87.62%, 85.67%, 84.08% and 69.66% respectively. Gao et al.[55] proposed a deep learning algorithm to learn its features and classify the cognitive states of emotion.

Different from conventional methods, the proposed algorithm perform hierarchical feature learning with 3-layer of Restricted of Boltzmann Machines (RBMs) on raw EEG signals without explicit handcrafted feature extraction. It provided 4-state emotion classification with 68.4% accuracy. Yang et al. [56] presented a hierarchical network with subnetwork nodes for emotion classification. The top layer of the network combines all features derived from subnetwork nodes and perform a re-mapping these features to provide more reliable classification. This work achieved 91.36% accuracy for 3-state valence classification.

Chapter 4
Static DLN Approach

4.1 Research Methodology

4.1.1 System Architecture

The EEG-based emotion classification system is developed with a stack of (denoising) autoencoders and two softmax classifiers on top of the network. The system will be able to perform emotion classification by determining valence status and arousal status simultaneously. The system architecture is shown in Figure 15. First, the system obtains EEG signals from an amplifier and then the signals are pre-processed by band-pass and notch filters. Input features of the DLN is a set of EEG power spectral values on each channel, extracted in five frequency bands i.e. Theta, Low Alpha, High Alpha, Beta and Gamma. Subsequently, valence and arousal states are used for emotion classification.
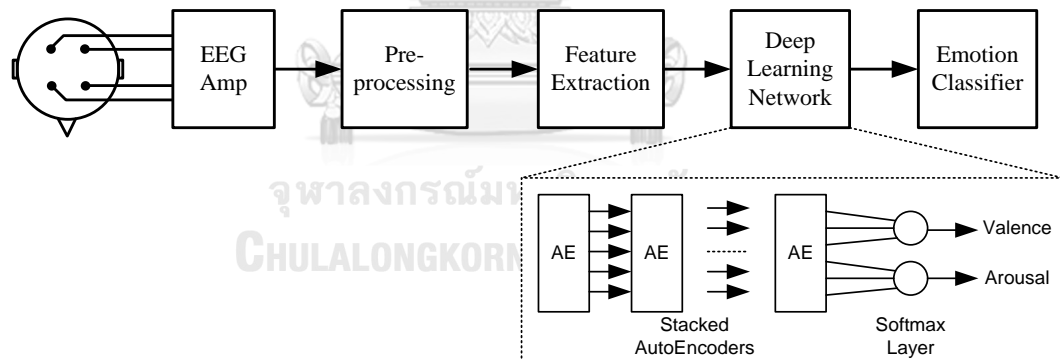
Figure 15   Static DLN System Architecture

4.1.2 Stacked Denoising Autoencoder

A stack of autoencoder is a special kind of neural networks. Figure 16 illustrates the implementation structure of an autoencoder. The autoencoder is to learn an identify function, shown in Equation (1).

$$\widehat{x} = h_{W,b}(x) \approx x \tag{1}$$

where $\widehat{x}$: Approximated outputs
x : Raw inputs
h : Learned identify function
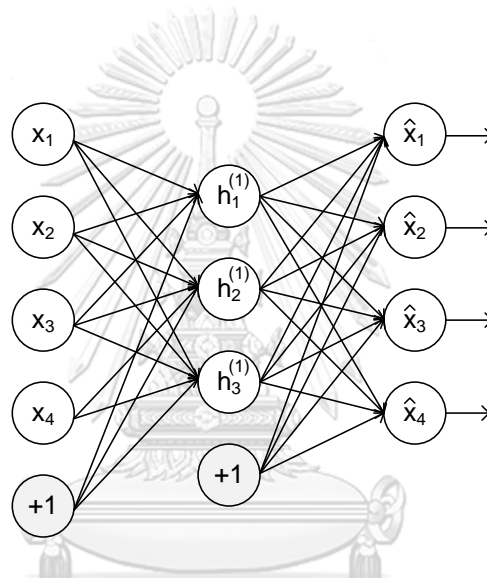
Figure 16   Implementation Structure of autoencoders

DLNs provide an unique greedy layer-wise unsupervised pre-training algorithm, beginning from its inputs to its outputs. The first layer of autoencoder learns on its inputs to investigate unknown feature coherences among them. During pre-training, all weight and bias paramters of the network will be adjusted to minimize its cost function, shown in Equation (2). Subsequently, the stack of autoencoders performs forward propagation to obtain the primary features of its inputs. To perform unsupervised pre-training in the next layer, the stack of autoencoders calculates its features in the same method by using features from the previous layer.

$$\text{Cost} = \frac{1}{2n} \sum_{i=1}^{n} (\hat{x}_i - x_i)^2 + \beta \sum_{j=1}^{m} KL(\rho || \hat{\rho}_j) + \frac{\lambda}{2} \sum_{i=1}^{n} \sum_{j=1}^{m} \theta_{ij}^2 \qquad (2)$$

where   m: Number of hidden nodes
          n: Number of inputs
          β: Weight of sparsity penalty
          KL: Kullback-Leibler divergence function
          ρ: Sparsity parameter
          $\hat{\rho}_i$: Probability of firing activity
          λ: Weight decay parameter
          θ: Weight of hidden nodes

The reconstruction criterion in the autoencoder model can not guarantee the extraction of useful features. We need to modify the reconstruction criterion for cleaning partially corrupted input with denoising approach. This approach allows the algorithm to make a higher level representation more stable and robust under corruptions of the inputs. A denoising autoencorder attempts to reconstruct a clean input from its corrupt version. First, we partially corrupt the initial input x into $\tilde{x}$ by means of a stochastic mapping (qD). Subsequently, the corrupted verion of its input $\tilde{x}$ is mapped to $y = f_\theta(\tilde{x})$ from which the algorithm attempts to reconstruct $z = g_{\theta'}(y)$. Figure 17 [57] illustrates the schematic representation of the denoising autoencoder. Throughout the algorithm, parameters θ and θ' are trained to minimize its average reconstruction error which is a squared error over a training set.
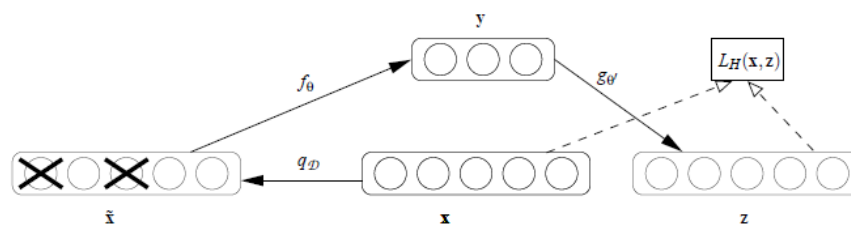
Figure 17   Denoising autoencorder schematic

4.1.3 Softmax Classifier

The main responsibility of softmax classifiers is to statistically estimate the probability of output features of deep learning networks. Softmax classifiers learn to adjust all values of weights and biases by using features of the highest layer. A stack of autoencoders with a softmax classifier is shown in Figure 18. In binary classification (k=2), the softmax regression hypothesis outputs $h_\theta(x)$ is shown in Equation (3).

$$h_\theta(x) = \frac{1}{e^{\theta_1^T x} + e^{\theta_2^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x} \\ e^{\theta_2^T x} \end{bmatrix} \tag{3}$$

In case of multi-class classification, the hypothesis of a softmax classifier is a vector of k estimated probabilities, shown in Equation (4).

$$h_\theta(x) = \frac{1}{\sum_{j=1}^{k} e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \vdots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix} \tag{4}$$
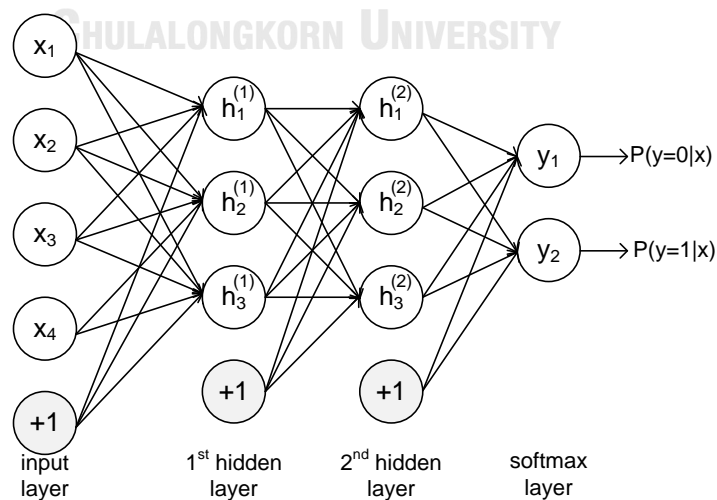
Figure 18   Autoencoder stack with softmax

The algorithm uses supervised learning approach to learn the weight and bias parameters by minimizing its cost function, shown in Equation (5).

$$\text{Cost} = -\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{k} 1\{y_i=j\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^{k} e^{\theta_l^T x^{(i)}}} + \frac{\lambda}{2}\sum_{i=1}^{k}\sum_{j=1}^{n} \theta_{ij}^2 \tag{5}$$

where   m: Number of hidden units
n: Number of inputs
k: Number of classes
y: Ground truth
θ: Weight of hidden nodes

After supervised training of the softmax layer finishes, the DLN has to perform its fine-tuning procedure to adjust all of weights and biases in the entire network by using backpropagation. The conventional backpropagation is utilized to learn all weights and biases of the network based on labeled data samples. The effort of this training is to minimize classification errors.

4.1.4   EEG Dataset

In static DLN approach, we used DEAP dataset [58] which is a multimodality dataset for emotion state estimation. EEG signals and other peripheral physiological signals of this dataset were collected from 32 subjects while each subject was watching 40 of one-minute videos. The subjects evaluted their levels of valence, arousal and dominance, after watching each music video. They used Self-assessment manikins (SAM) [59] to show the scales of valence and arousal states as shown in Figure 19. The subjects selected a scale numbers 1-9 in order to determine their affective states in each categories.
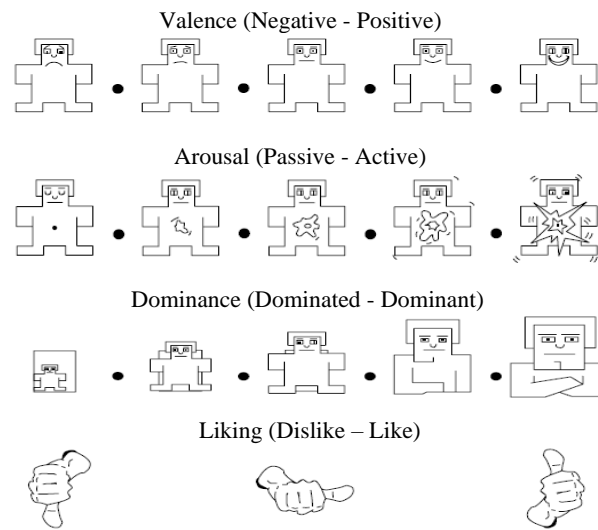
Figure 19   SAM for emotion state assessments

From self-assessment values, we mapped the scale numbers (1-9) into three levels of each valence and arousal states. For valence states, the scale numbers 1-3 were mapped to "Negative", the scale number 4-6 to "Neutral" and the scale number 7-9 to "Positive", respectively. For arousal states, the scale numbers 1-3 were mapped to "Passive", the scale number 4-6 to "Neutral" and the scale number 7-9 to "Active", respectively. According to the scale mapping, the system provides 9-category states of emotions: Happy, Pleased, Relaxed, Excited, Neutral, Calm, Distressed, Miserable and Depressed as shown in Figure 20.
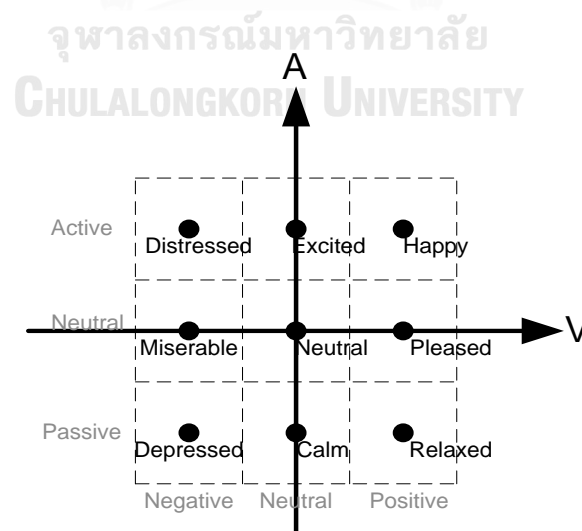
Figure 20   Cognitive states of emotions

4.1.5 EEG Feature Extraction

Our proposed system acquired all 32 channels of EEG signals in the DEAP dataset. Other peripheral physiological signals were not been used in this research study. The EEG signals were down sampled from 512 Hz into 128 Hz. The EEG power spectral was computed using a Matlab FFT function with Hanning window size of 128 samples. The EEG power spectral features of all 32 channels were extracted into five ranges of power bands: Theta (4-8Hz), Low Alpha (8-10Hz), High Alpha (10-12Hz), Beta (12-30Hz) and Gamma (30Hz up). Furthermore, the different EEG power features of all symmetrical 14 pairs on the right and the left hemispheres were extracted. The purpose of these features is to measure the possible asymmetry in brain activities due to emotional stimuli. A total number of 230 EEG features were used as feature inputs of the DLN.

All of these EEG power spectral features were substracted with their baseline power to provide the change of power relative to their pre-stimulus period. Subsequently, the EEG power spectral features were normalized into the range (0.1, 0.9). This normalization process is neccessary since our implementation uses sigmoid as the activation function in the output layer. Some of the features outside the range of -3*SD and +3*SD were truncated into 0.1 and 0.9 respectively.

4.1.6 EEG-based Emotion Classification with DLNs

Our proposed system is developed with a stack of three autoencoder layers and two softmax layers on top of the network as shown in Figure 21. The system estimates the cognitive states of valence and arousal emotions separately. The outcome of unsupervised pre-training procedure can be shared between two softmax layers, one for valence and another for arousal, because both softmax classifiers employ the same set of unlabeled data. However, two softmax classifiers need to use their own stacked autoencoders during fine-tuning backpropagation.
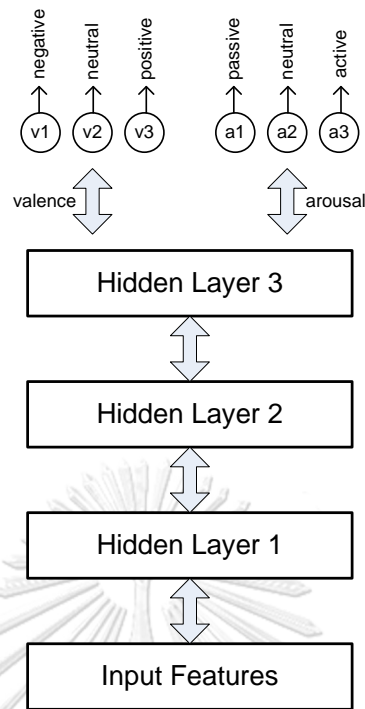
Figure 21   DLN with two softmax classifiers

DLNs utilize the unsupervised pre-training technique with greedy layer-wise training. The first sparse autoencoder is trained on the 230 features in order to learn the primary features on these input features. We employ the L-BFGS algorithm to minimize the cost function which is the sum of squared errors between input features and output values. Table 2 shows all of configuration parameters in our proposed system. Next, the network performs forward propagation from its input features to the sparse autoencoder layer to obtain primary features. These features must be used to perform unsupervised pre-training in the second hidden layer. The algorithm continues to learn its features in the same way until the last hidden layer.

Table 2   DLN Parameter Settings

Parameters	Value
MaxIteration - SAE	400
MaxIteration - Softmax	100
MaxIteration - Finetuning	200
Hidden Layer	100, 50
Sparse Parameter (ρ)	0.10
Weight of sparse penalty (β)	3.0
Weight Decay Parameter (λ) - SAE	0.003
Weight Decay Parameter (λ) - Softmax	0.0001
Weight Decay Parameter (λ) - Finetuning	0.001

With supervised learning approach, the weight and bias parameters of two softmax classifers are properly trained by using the output features of the last hidden layer of the DLN. A set of self-assessment emotion states in the DEAP dataset is used as labels. These two softmax classifers can learn their own parameters concurrently. After the algorithm completes learning both softmax's parameters, the algorithm needs to perform fine-tuning procedure on all parameters in the entire network simultaneously. However, it is required to save the network parameters deriving from unsupervised pre-training process and then re-assign these parameters for fine-tuning process of another softmax classifier. The fine-tuning process treats the entire network as a single model and then adjusts all parameters with backpropagation. The backpropagation process is used to learn the weight and bias parameters based on labeled training datasets to minimize classification errors.

The DLN training protocol [11] is explained in Figure 22. The greedy layer-wise unsupervised pre-training algorithm begin from the first hidden layer to the last hidden layer. Initial weight and bias parameters of each hidden layer are assigned for parameter optimizations. Subsequently, the features deriving from feedforward propagation at each hidden layer must be used to perform unsupervised pre-training in the next hidden layer. The process will continue the same procedure until the last hidden layer of the network. After the unsupervised pre-training in the last hidden layer finishes, softmax training and fine-tuning procedures need to be operated respectively.
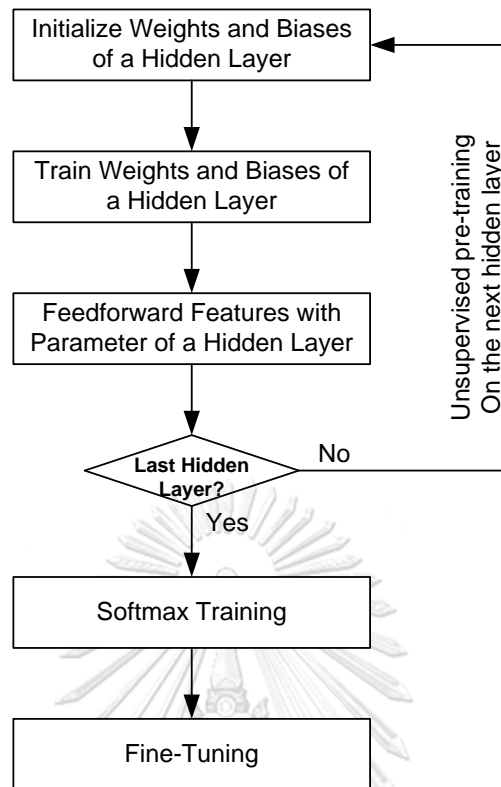
Figure 22   DLN training protocol

4.1.7 Covariate Shift Adaptation Of Principal Component

Our proposed EEG-based emotion classification system, developed with a stack of autoencoders, has potential to represent a highly expressive abstraction. Therefore, the system has a chance to confront with overfitting problems, especially with a large number of hidden nodes and input features. Furthermore, EEG signals have a non-stationary effect and it is difficult to develop a reliable classification. Our proposed system applies the concept of principal component based covariate-shift adaptation [60] to alleviate the problems of overfitting and non-stationary effects simultaneously. Principal Component Analysis (PCA) [61] is used to extract the most significant feature and normalize these features by shifting a window over the data to alleviate the effect of non-stationarity.

4.2 Experimental Results

This research study evaluated the effectiveness of our EEG-based emotion classification system with four experiment configurations as shown in Figure 23. All experiment setups utilized a deep network with a stack of three autoencorders and two softmax classifiers. In the first experiment, we used one hundred hidden nodes in each layer, called "DLN-100". To investigate the impact of the size of hidden neural nodes, we reduced the number of hidden nodes to fifty nodes, called "DLN-50". The third experiment applied PCA algorithm to reduce the dimensionality of feature inputs from 230 to 50 features in order to alleviate overfitting problem. The size-reduced features were inputs to the stack of autoencoders with fifty hidden nodes in each layer. Finally, we exploited the concept of covariate shift adaptation (CSA) to relieve the non-stationary effects of EEG signals. This experiment performed a normalization of its feature inputs with the average values of the features within a window equal to 10.
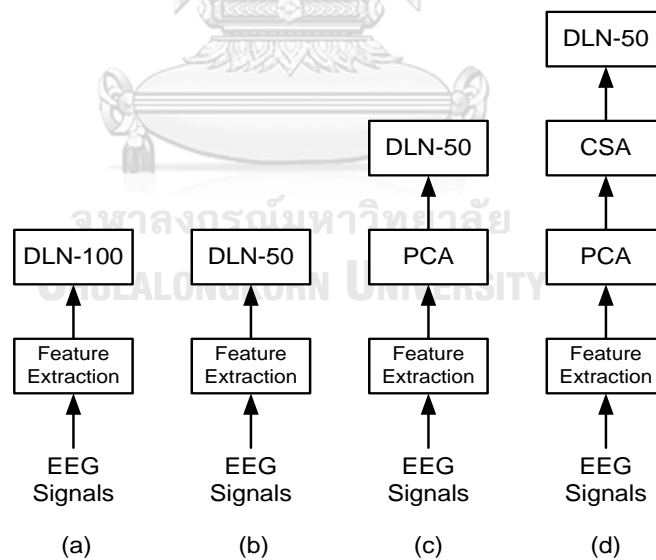
Figure 23   Overview of Experiment Setups

In this static approach, the accuracy performance of valence and arousal emotion classification in all experiment setups was evaluated with a leave-one-out cross-validation scheme. The experimental results of these experiment configurations were shown in Table 3. Moreover, the average values of classification accuracy for individual subjects were illustrated in Figure 24. The DLN-100 configuration provides the valence accuracy of 49.52% and the arousal accuracy of 46.03%. Those of DLN-50 configuration slightly decreased into 47.87% for valence and 45.50% for arousal respectively. The size of autoencoder layers in our experiment setups affected the accuracy performance of the cognitive states of emotions. From experimental results, when the number of hidden nodes in each layer was reduced from 100 to 50 nodes, the classification accuracy decreased 1.62% and 0.53% for valence and arousal states respectively. In the third experiment setup, the PCA configuration enhanced the classification accuracy of valence by 3.01% and those of arousal by 3.14%. Finally, the system configuration with the CSA provides 53.42% accuracy for valence state classification and 52.03% accuracy for arousal state classification. We found that the PCA+CSA configuration was able to enhance the accuracy by 5.55% for valence and 6.53% for arousal.

This research study fully investigated the effectiveness of a deep learning network for EEG-based emotion classification by comparing its accuracy performance with a SVM classifier. We utilized LIBSVM tools [63] to perform evaluation of accuracy performance for the SVM classifier by setting its kernel function to radial basis function. Other parameters of SVM configuration were assigned by default values. We established three experiment configurations for the SVM classifier: 230 input features, PCA and PCA+CSA. The accuracy performance of the SVM classifier on these configurations is shown in Table 3.

Table 3   Accuracy for Valence and Arousal Classification

	Valence Accuracy (%)							Arousal Accuracy (%)						
	DLN				SVM			DLN				SVM		
	N100	**N50**	**PCA**	**CSA**	**F230**	**PCA**	**CSA**	**N100**	**N50**	**PCA**	**CSA**	**F230**	**PCA**	**CSA**
S01	46.14	44.64	50.31	**56.52**	38.63	40.29	42.92	**50.21**	48.98	38.57	42.85	41.67	42.58	23.83
S02	47.64	44.72	**49.73**	48.48	41.50	44.88	47.04	35.85	**37.56**	34.65	36.43	32.62	23.25	17.46
S03	52.45	51.43	58.10	**59.81**	47.38	51.75	40.17	42.84	40.61	42.86	**54.43**	31.63	22.08	12.38
S04	39.20	36.43	37.52	**40.98**	24.21	22.12	29.92	39.45	36.98	**52.20**	46.22	16.50	22.58	34.04
S05	54.32	56.76	55.56	**60.97**	44.58	33.33	49.42	48.62	47.31	**58.78**	53.10	41.75	39.13	47.00
S06	49.48	47.43	54.81	**71.06**	44.50	50.21	60.08	49.24	49.77	45.65	**62.18**	43.46	29.71	20.29
S07	51.87	52.51	57.94	**73.48**	47.00	50.58	42.88	46.25	44.27	53.15	**56.01**	41.13	37.79	39.83
S08	**49.19**	48.81	54.27	48.44	39.58	47.08	37.54	54.44	52.19	57.24	**60.14**	46.92	45.25	52.71
S09	55.86	**58.81**	63.73	38.48	43.00	36.96	27.50	49.81	49.06	62.15	**64.18**	47.88	51.25	54.46
S10	43.54	40.43	**45.27**	40.98	39.47	32.83	30.00	41.21	39.52	52.40	**64.16**	34.96	34.38	25.88
S11	43.87	40.31	44.90	**48.11**	34.04	37.46	37.38	35.68	34.48	35.45	**35.85**	23.58	27.54	25.08
S12	44.32	41.56	43.40	**48.06**	37.79	39.21	42.12	50.74	49.56	44.45	50.81	**51.25**	46.96	36.88
S13	**54.86**	53.31	44.10	47.36	49.38	32.92	30.13	**48.65**	48.31	38.07	41.26	35.08	27.54	28.17
S14	33.81	35.64	43.69	**44.06**	30.00	35.58	44.67	51.96	49.27	61.99	**62.14**	44.67	51.63	42.13
S15	**58.74**	57.45	42.90	46.60	52.13	40.63	40.00	48.55	47.90	64.15	**65.01**	36.29	29.46	23.25
S16	**47.95**	45.76	35.69	38.77	36.92	29.83	25.00	41.29	40.61	**50.98**	49.64	39.33	39.67	21.88
S17	**53.20**	49.35	45.56	47.86	51.50	44.56	40.00	56.58	58.98	61.61	**62.89**	50.79	38.83	33.54
S18	55.21	53.72	56.65	**57.98**	40.92	43.63	42.50	51.43	54.73	62.82	**66.47**	51.54	39.58	54.33
S19	56.38	53.51	55.90	**62.27**	39.04	45.38	40.42	47.19	47.52	49.74	**59.81**	46.67	38.00	42.67
S20	48.65	46.31	62.85	**65.11**	51.29	42.04	40.00	52.63	**56.73**	54.95	55.26	50.38	34.79	37.17
S21	51.78	48.14	67.98	**70.23**	46.92	46.71	42.50	**45.97**	44.27	37.15	41.85	40.75	35.83	25.58
S22	42.97	43.22	40.44	**45.56**	37.46	31.71	37.71	47.11	45.65	49.40	**52.60**	36.13	43.79	30.79
S23	58.43	55.01	58.73	**61.73**	48.75	48.58	47.50	28.45	31.02	31.15	**36.35**	24.88	23.25	22.50
S24	**49.74**	46.81	45.27	45.73	36.54	40.63	35.00	59.45	61.15	58.32	**62.72**	60.04	46.46	39.67
S25	35.72	36.26	41.56	**45.06**	31.33	26.04	37.50	**40.74**	40.65	37.49	39.68	33.33	24.54	25.13
S26	43.16	40.51	45.85	**52.90**	36.33	36.58	28.21	41.88	39.27	45.07	**48.10**	27.70	23.54	29.38
S27	58.65	**60.14**	49.94	52.61	40.58	56.17	54.88	45.86	**46.31**	42.03	42.81	44.96	34.29	37.13
S28	48.85	46.76	49.35	**53.77**	38.13	35.79	46.83	40.22	39.52	40.45	**45.32**	31.71	32.33	24.79
S29	51.25	48.06	51.23	**54.52**	41.79	45.04	49.00	35.44	35.15	37.28	**38.60**	27.33	35.21	19.46
S30	**56.40**	53.76	52.81	54.36	40.46	29.00	34.63	52.21	49.23	41.49	**54.64**	33.79	32.08	32.33
S31	51.34	48.10	59.02	**61.69**	43.04	42.46	44.71	41.76	40.02	**55.20**	51.97	34.67	29.79	36.13
S32	49.67	46.06	63.19	**65.90**	41.75	34.63	40.00	51.28	49.36	59.49	**61.39**	45.25	43.38	42.83
Mean	*49.52*	*47.87*	*50.88*	*53.42*	*41.12*	*39.83*	*40.26*	*46.03*	*45.50*	*48.64*	*52.03*	*39.02*	*35.21*	*32.46*
S.D.	*±6.34*	*±6.57*	*±8.18*	*±9.64*	*±6.39*	*±7.94*	*±7.87*	*±6.84*	*±7.17*	*±9.85*	*±9.74*	*±9.59*	*±8.56*	*±10.90*
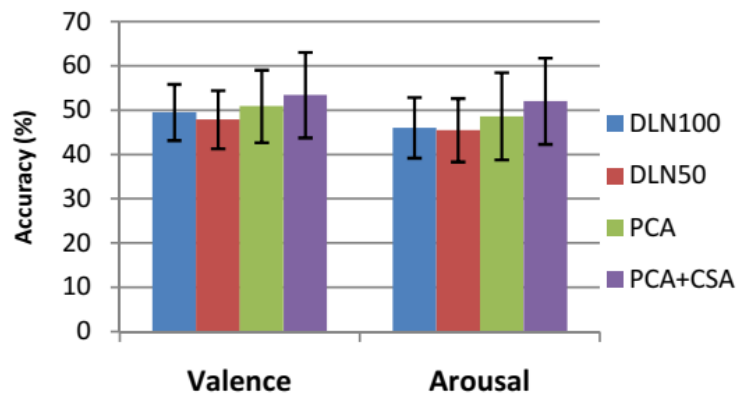
Figure 24   Average accuracy of the experiments

4.3 Discussion

The number of hidden nodes in the DLN effected on the accuracy performance of our proposed emotion classification. We found that the more number of hidden neural nodes to perform hierarchical feature learning provided better classification accuracy results. From experimental results in previous section, the DLN-100 configuration has better accuracy performance compared of the DLN-50 configuration. For valence state classification, the accuracy of the DLN-100 is 49.52% but those of the DLN-50 is 47.87%. For arousal state classification, the accuracy of the DLN-100 is 46.03% but those of the DLN-50 is 45.50%. The DLN-100 configuration slightly increased its accuracy compared of the DLN-50.

The concept of Principal Component Analysis (PCA) provides a helpful mechanism for the deep architecture to learn linear correlations between feature inputs. It performs as one additional layer of the network to boost the performance of the learning task. There is a relationship between autoencoder and PCA [62]. In our experiment configuration, the number of hidden nodes is less than the number of input nodes. The autoencoder essentially performs Nonlinear Principal Analysis (NPCA). If there are some unknown correlations among the inputs in both linear and non-linear correlations, then these two algorithms will be able to discover some of those correlations. From experimental results, the PCA increases the accuracy performance by 3.01% for valence and 3.14% for arousal.

The benchmark of our proposed system with a SVM classifier is to compare the efficiency of accuracy performance. Figure 25 and Figure 26 illustrates the comparison between DLN and SVM classification for valence and arousal states respectively. We found that the DLN provides better accuracy performance compared with the SVM in all experiment setups. Moreover, the PCA was able to enhance the accuracy performance in the DLN but it provides lower accuracy results in the SVM. The effect of PCA on SVM is congruent with a study by K. Li et al [64].
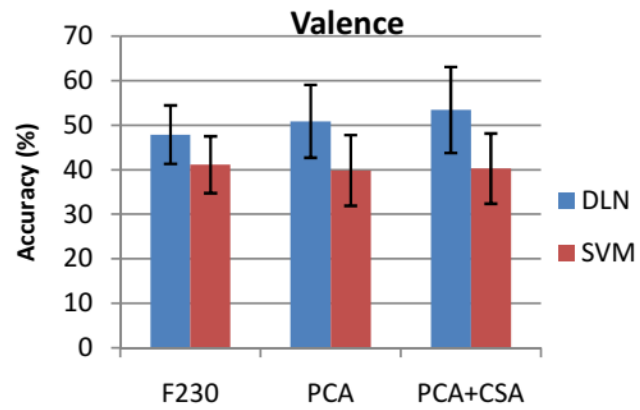
Figure 25   DLN vs SVM Accuracy for Valence
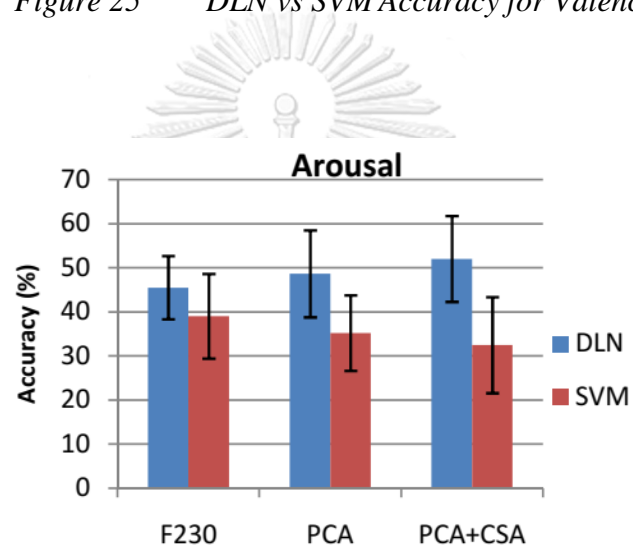
Figure 26   DLN vs SVM Accuracy for Arousal

Chapter 5
Dynamic DLN-HMM Approach

5.1 Research Methodology

5.1.1 System Architecture

The new proposed EEG-based emotion classification system utilizes Hidden Markov Models (HMMs) to learn meaningful state transitions of high-level features deriving from Deep Learning Networks (DLNs). The overview of system architecture of EEG-based emotion classification with dynamic approach is shown in Figure 27. The HMM is responsible of learning the cognitive states of emotions by investigating the characteristic changes of the DLN's outputs. First, a series of T-window inputs are fed into m-layer of the DLN. The DLN provides hierarchical feature learning method. Features at its high-level can be learnt from compositions of features at its low-level with greedy layer-wise unsupervised pre-training. Subsequently, the n-state HMMs examines the transition of top-level features and then determines the most probability of emotions.
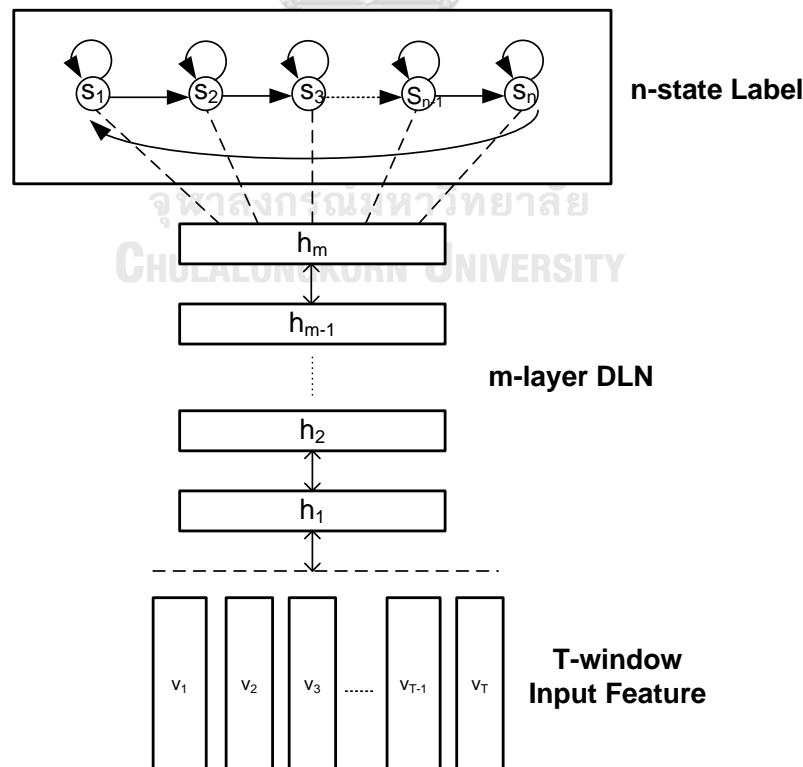
Figure 27   System Architecture with Hidden Markov Models

5.1.2 Hybrid DLN-HMM Method

The proposed system uses one HMM model representing each valence/arousal state. A hybrid DLN-HMM system, in Figure 28, uses three HMM models that represent levels of valence status (Positive, Neutral and Negative) and each HMM model has three states to learn meaningful state transitions of high-level features of the DLN. During training sessions, the DLN attempts to map each sequence of input features to a desired output state, representing as a target softmax node. During classifying, the system utilizes the value of output probability of each softmax node to determine which HMM model (valence status) has the maximum likelihood among them.
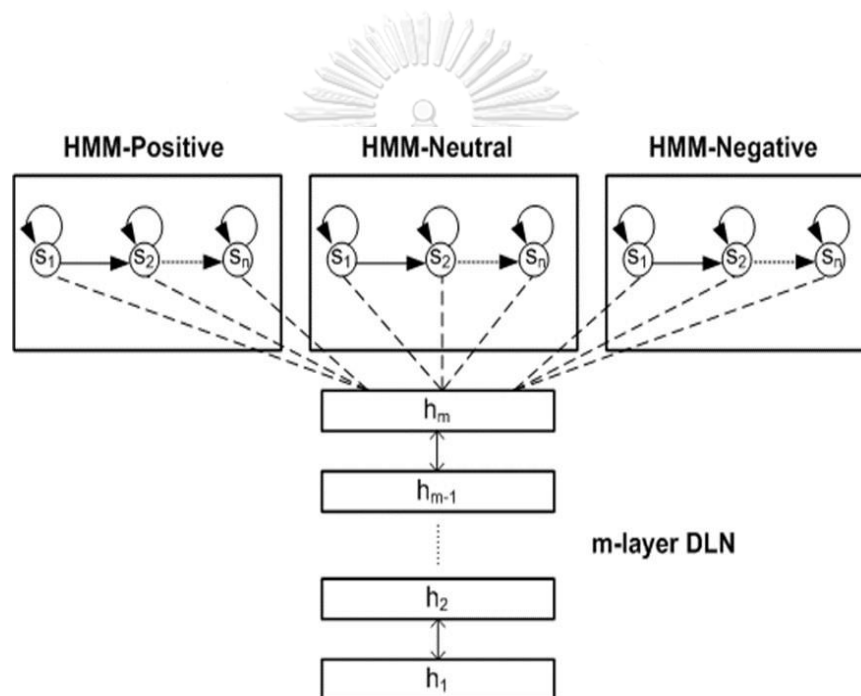
Figure 28   Hybrid DLN-HMM model

5.1.3 EEG Dataset

This research study examines the emotion state transitions by investigating the meaningful changes of high-level features of Deep Learning Networks (DLNs) with Hidden Markov Models (HMMs). In order to investigate the temporal neural dynamics of different emotions, a new emotion stimulus experiment was established to collect EEG data. Our required dataset specification is different from other existing available datasets. In this research study, we focus on learning state transitions of HMMs in short period (20 seconds) with a single stimulu event. Therefore, each data sample in our dataset consists of a sequence of EEG features in 20 seconds.

The DEAP dataset [58] provides a multimodality dataset for emotion state classification. EEG signals were collected from 32 subjects while each subject was watching 40 of one-minute videos. This means the dataset provides only 40 data samples of valence and arousal classification for each individual subject. Our system evalutation is based on subject-dependent protocol because EEG signals have the characteristics of high non-stationarity and inter-subject variability. Therefore, there are approximately 10-15 data samples for each valence/arousal states. Such a small number of data samples for each emotion state is insufficient to employ the DEAP dataset for our proposed methodology.

Our emotion stimulus protocol is depicted in Figure 29. The system starts with a signal on the monitor to perform data frame synchronization. The participants need to blink their eyes three times with signals on the monitor. After that the first picture is displayed on the monitor for 20 second and then blank screen for 20 second. In each session, the system displays a set of five pictures. Our research study utilizes a set of pictures in International Affective Picture System (IAPS) with background music for emotion stimulation. Each participant needs to perform emotion stimulus for recording 200 samples of each emotion (Happy, Relax, Sad, Afraid). We used the blank picture to collect raw EEG signals for neutral emotion.

Before the experiment sessions began, we informed the participants about the emotion stimulus experiment and instructed them to sit comfortably watching all upcoming pictures on the screen. We emphasized the participants to watch the pictures attentively and refrain from any movement as much as possible. We could asked the participants to re-start a new recording session again, if any major movement occurred.
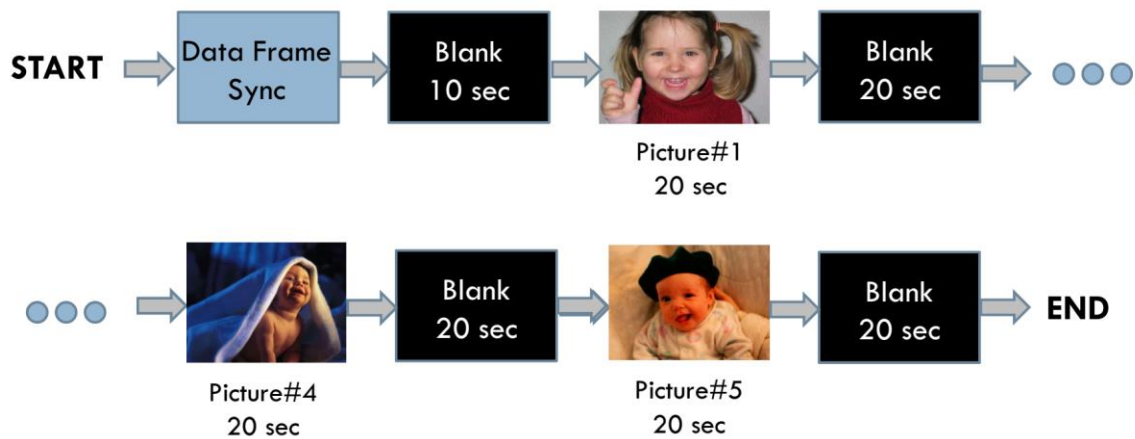
Figure 29   Emotion Stimulus Protocol

Raw EEG signals were acquired from a research-grade EEG amplifier device, called "g.Nautilus". The EEG amplifier has 16 channel (FP1, FP2, F3, Fz, F4, C3, Cz, C4, T7, T8, P3, Pz, P4, PO7, PO8 and Oz) with 24-bit accuracy at sampling rate 250Hz. During EEG signal recording, subjects comfortably sat on a chair in front of 23" LED monitor with 80-100 cm distance, shown in Figure 30.

Figure 30   Emotion Stimulus Session.

Each participant has total number of 200 data samples of each emotion – Happy, Relax, Sad and Afraid. Also, there are 400 data samples of Neutral emotion from blank picture stimulus. For valence classification, the experiment uses its positive valence data samples from Happy and Relax emotion, and its negative valence data from Sad and Afraid, as well as its neutral valence data from neutral emotion stimulus. For arousal classification, we use the active arousal data from Happy and Afraid, and passive arousal data from Relax and Sad, as well as neutral arousal from neutral emotion stimulus.

5.1.4 Feature Extraction

EEG feature extraction procedure is illustrated in Figure 31. First, each raw EEG signal record, derived from each stimulus session (5-picture stimulus), need to be separated into one-picture data frame. Then, each data frame (20-second period) was required to perform EEG artifact removal using Independent Component Analysis (ICA) algorithm. Subsequently, power spectral density was calculated using FFT with a Hanning window of size 250 samples. The window moves forward to 50 samples for next feature extraction. Therefore, each one-picture data frame (20 seconds) generates a sequence of 98 dataset, shown in Figure 32. The power spectral features of EEG signals on these channels were extracted in 5 frequency bands: Theta (4-8Hz), Low Alpha (8-10Hz), High Alpha (10-12Hz), Low Beta (12-16Hz) and High Beta (16-30Hz). In summary, each data sample consists of 96 sequence of 80 power spectrum features.
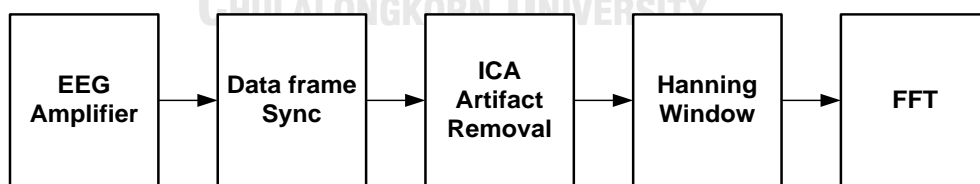
Figure 31   EEG feature extraction procedure
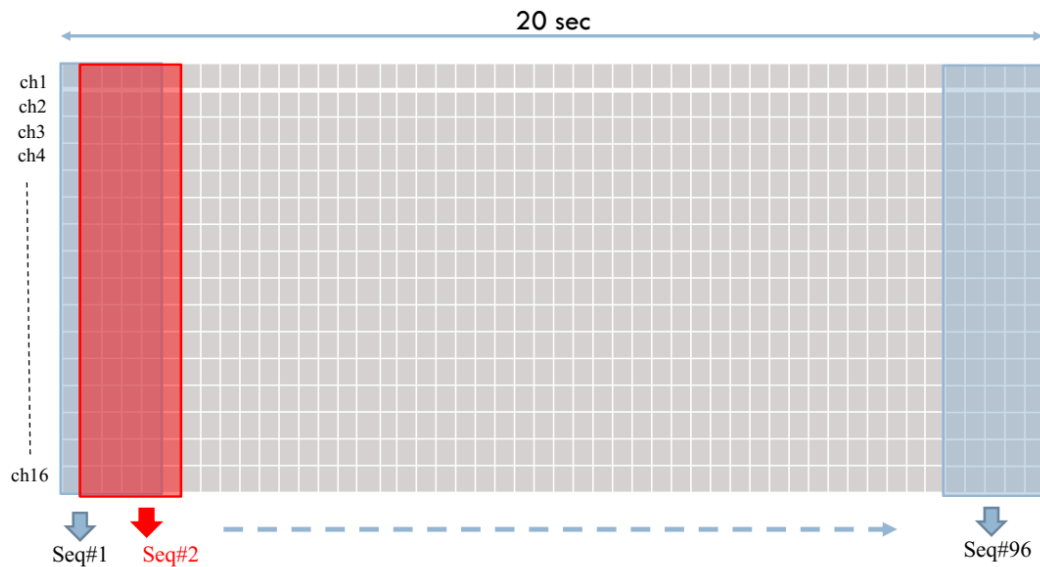
Figure 32   Feature Extraction Windowing

5.1.5   DLN-HMM Training Procedure

The EEG-based emotion classification with dynamic approach exploits the advantage of Hidden Markov Model (HMM) to learn relevant transitions of high-level features of Deep Learning Network (DLN). The softmax output nodes of the DLN are represented as individual state of the HMMs. Each valence status has its own HMM model – Positive, Neutral and Negative. The DLN-HMM training procedure is constituted of three steps: Gaussian Mixture Model (GMM)-HMM implementation, Optimal State Sequence Mapping and Label State Training, as shown in Figure 33.
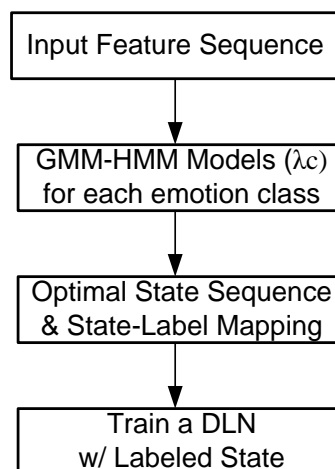
Figure 33   DLN-HMM Training Procedure

In the first step, the proposed methodology needs to generate an individual GMM-HMM of each valence status-- Positive, Neutral and Negative. In addition, each arousal status has its own GMM-HMM model-- Active, Neutral and Passive. Each GMM is a statistical model for EEG power spectral features of each emotion state. The property of Gaussian mixture distribution is its multi-modality, in contrast to the uni-modality of the Gaussian distribution (one component). This key advantage of the mixtures of Gaussian distribution is to adequately describe the prominent properties of complicated data signals, such as EEG signals. The multimodality in EEG data derives from multiple brain wave activities or events and each event is responsible for one component in the distribution. It is potentially identify each individual event and then the mixture distribution is the combination of such events of each emotion.

If a variable x has a Gaussian mixture distribution, its probability density function will be a form of

$$p(x) = \sum_{m=1}^{M} \frac{c_m}{(2\pi)^{1/2} \sigma_m} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_m}{\sigma_m}\right)^2\right] \qquad (6)$$

$$\text{where} \quad c_m > 0 \text{ and } \sum_{m=1}^{M} c_m = 1$$

In case of normal random vector $x = (x_1, x_2, ..., x_D)^T$, the mixture Gaussian distribution has the joint probability density function in form of

$$p(x) = \sum_{m=1}^{M} \frac{c_m}{(2\pi)^{D/2} |\Sigma_m|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu_m)^T \Sigma_m^{-1}(x-\mu_m)\right] \qquad (7)$$

$$\text{where} \quad c_m > 0 \text{ and } \sum_{m=1}^{M} c_m = 1$$

In this research study, the dimension of input feature x is large (D=80), then the use of full covariance matrix will use a large number of parameters. To reduce the number of parameters, the algorithm uses diagonal covariance matrix to generate GMM models. In additional, when using diagonal covariance matrix, the complexity of computation significantly reduces. The simplification of computation by implementing diagonal covariance matrix instead of full covariance matrix seems to cause uncorrelation among input vector components. However, each component of the Gaussian mixture has a diagonal covariance matrix, which effectively presents the correlations among input vector components.

Subsequently, HMM is responsible for learning relevant transitions in data distribution. This research study uses HMMlearn Toolkit version 0.2.1 [68] for generating GMM-HMM models of cognitive states of emotions, shown in Figure 34. The algorithm implements a circular left-to-right HMM for each valence/arousal status.
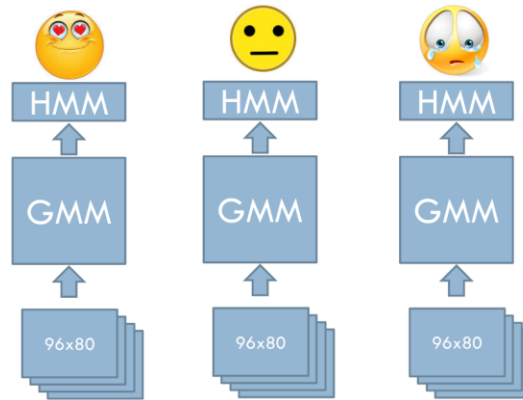
Figure 34   GMM-HMM model

The second step of DLN-HMM training procedure is to find optimal sequences of each data sample sequence in the training set, shown in Figure 35. The optimal sequence decoding is a process of finding the state sequence that have most likelihood probability. This research uses Viterbi decoding algorithm to find the optimal sequence of data sample sequence, as shown in Equation (8). Subsequently, each state in the optimal sequence is assigned a label L_i ($i \in \{1,...., CxQ\}$) by Q is number of states in HMMs and C in number of emotion classes.

$$P(O/\lambda c) = \max_{q1,...,q_T} \P_{q_1} \prod_{t=2}^{T} p(q_t|q_{t-1}) b_{q_t}(o_t) \qquad (8)$$

where λc: HMM model of emotion class (c)

O: Observation sequence (input sequence)

q: state of HMM

b : observation probability
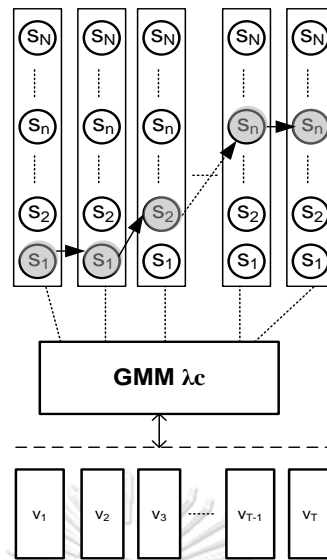
Figure 35   Optimal State Sequence Decoding

All training data sequences of all emotion classes need to be paired with its own labeled state sequence derived from optimal state decoding algorithm. All of these training data sequences are feature inputs to train a DLN. Its outputs are the posterior probabilities of CxQ units. This research study uses Keras Deep Learning Library [69], a high-level neural networks API with TensorFlow platform. The DLN is consisted of a stack of fully connected layer (Dense with "tanh" activation function) and one softmax layer on top, as illustrated in.Figure 36.
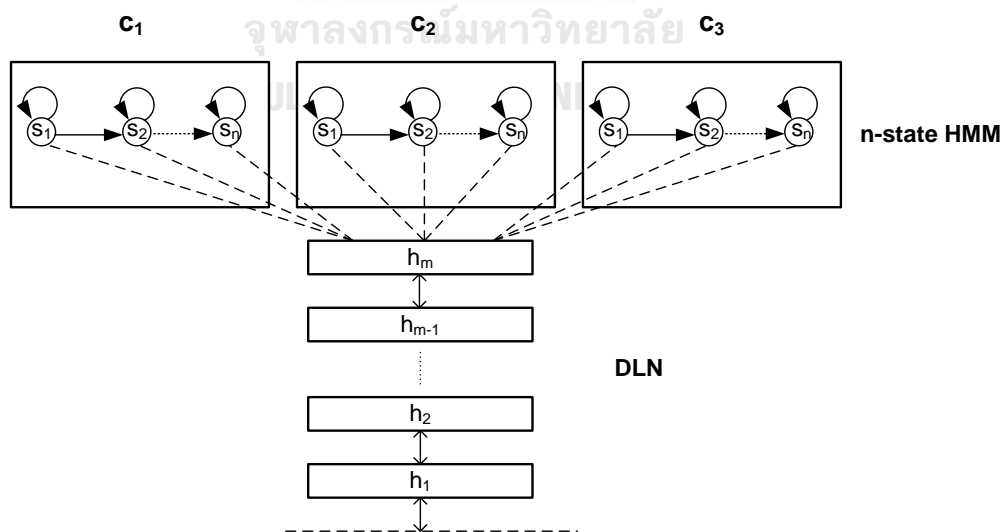
Figure 36   Softmax-layer and HMM-state mapping

5.1.6   DLN-HMM Classifying Procedure

After training all input feature sequences of all emotion classes, the DLN-HMM is capable of classifying emotion status of data samples in the testing set. The classifying procedure is constituted of Posterior Probability computation, state-label mapping, likelihood probability conversion and maximum likelihood computation, as shown in Figure 37.
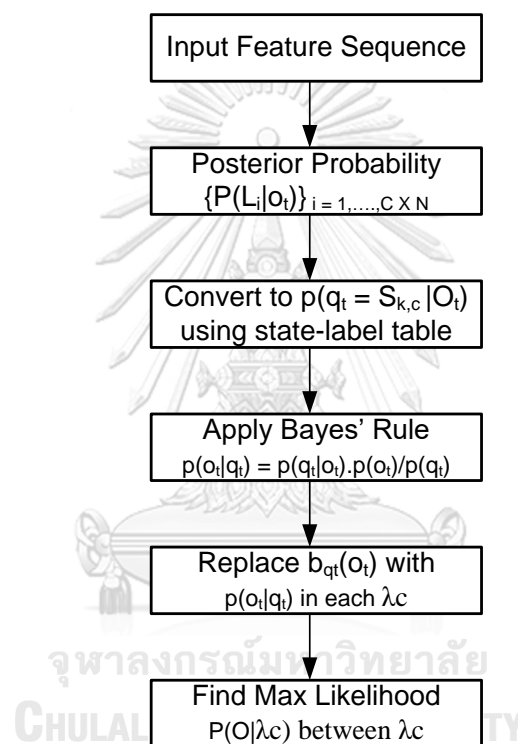
Figure 37   DLN-HMM Classifying Procedure

The first step of emotion classification is to feed the input feature sequence to the DLN. Posterior Probability of CxN state labels can be calculated from the DLN and then the algorithm performs state-label mapping to provide the posterior probability of each state in HMM models. The Posterior Probability computation process is illustrated in Figure 38.
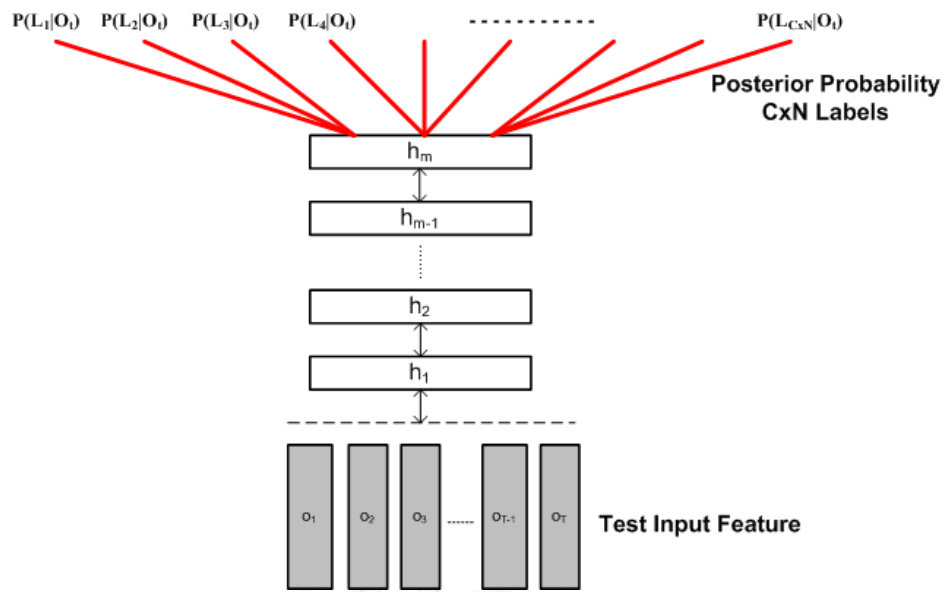
Figure 38   Posterior Probability Computation

To determine the maximum likelihood probability among emotion classes, the algorithm needs to convert the posterior probability $p(q_t|o_t)$ into the likelihood probability $p(o_t|q_t)$ by using Bayes's Rules, in Equation (9). In our proposed protocol, the prior probability of each state, $p(q_t)$, is computed as the probabity of state occupation. The probability $p(o_t)$ is assigned as a constant because of the independence of these feature inputs.

$$p(o_t|q_t) = \frac{p(q_t|o_t)p(o_t)}{p(q_t)} \tag{9}$$

Before performing Viterbi algorithm to calculate the likelihood probabilities $P(O|\lambda c)$. In HMM (λc) of each emotion class c, we replace the observation probability $b_{qt}(o_t)$ with $p(o_t|q_t)$. Finally, the algorithm determine the maximum likelihood probability among emotion classes, in Equation (10).

$$\text{Classification Result} = \text{argmax} \ \{P(O|\lambda_1), \ P(O|\lambda_2), \ P(O|\lambda_3)\} \tag{10}$$

5.2 Experimental Results

The main purpose of this dissertation is to develop an EEG-based emotion classification system with hybrid DLN-HMM implementation. The DLN is responsible for performing hierarchical feature learning of inputs. The HMM is to learn temporal dynamics of high-level features of the DLN. It examines all relevant state transitions of softmax output nodes in the DLN. To evaluate the DLN-HMM model, the system configuration, as shown in Figure 39, had been used. The model is required to complete the training procedure in Section 5.1.6, and then perform the classifying procedure in Section 5.1.7, respectively.
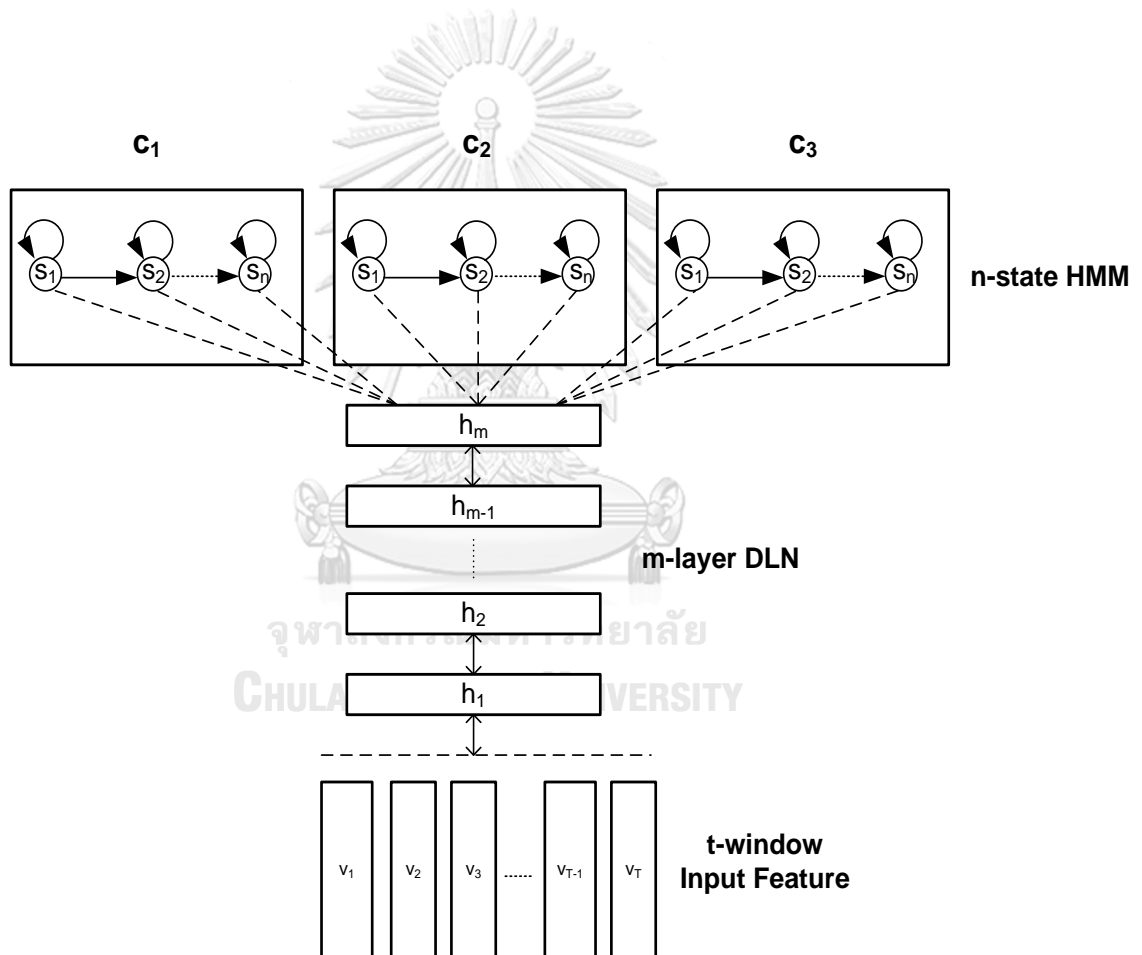
Figure 39   DLN-HMM Implementation

This experiment employs two DLN-HMM models to perform emotional states of valence and arousal separately, shown in Figure 40. Each emotion has its own DLN-HMM model implementation. However, these two models use the same configuration parameters. The reason of valence-arousal model separation is that the experiment requires each DLN-HMM performs intensive learning features and state transitions of its own cognitive state. In case, the experiment combines both valence and arousal state into one huge model. It is potentially harmful for its accuracy performance.
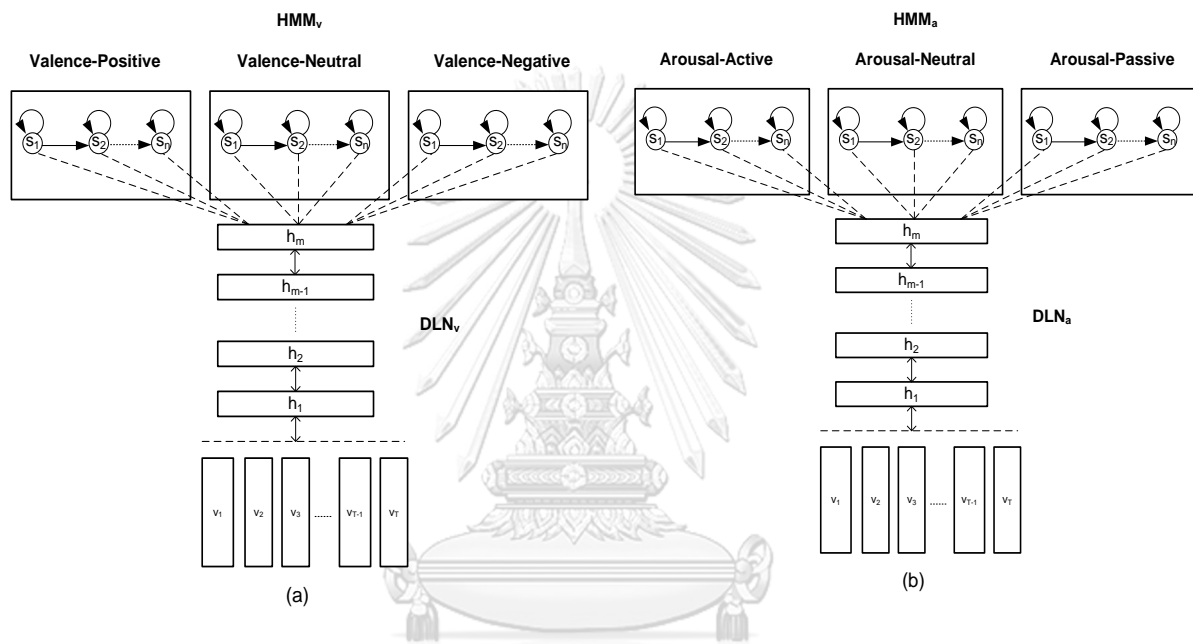
Figure 40   DLN-HMM implementation (a) Valence and (b) Arousal

This research study uses our own dataset derived from five subjects, age 24 to 44 year old. Each subject need to perform emotion stimulus for recording EEG 40 sessions of each emotion (Happy, Relax, Sad, Afraid and Neutral). In each session, the system displays a set of five pictures. Therefore, the dataset consists of 200 samples for each emotion. The detail of data recording protocol is described in Section 5.1.3.

This research study examines how the system's configuration parameters affect the emotion classification accuracy. These configuration parameters are the number of HMM states (n-state), the number of DLN hidden layers (m-layer) and the number of input feature frames (f-frame window). The experiment performs subject-dependent classification with 10-fold cross validation in each variant of system configuration parameters.

5.2.1   Variant of n-state HMM

The number of HMM states (n-state) is one of system configuration parameters affecting the DLN-HMM's accuracy performance. This section intensively investigates the variant of n-state HMM parameters by adjusting n = 3, 5 and 7 states respectively. All of other parameters have been fixed in each variant of n-state value. Table 4 lists all of parameters for variant of n-state HMMs. The system configuration is depicted in Figure 41.

Table 4   Configuration Parameters of n-state HMM Variant

Parameters		Value
Input	Number of input frame window(f-frame)	7 frames
	Number of Input Features	560
DLN	Number of DLN Layers (m-layer)	5 layers
	First Layer size	300
	Second Layer Size	200
	Third Layer Size	100
	Fourth Layer Size	50
	Fifth Layer Size	25
	Softmax Layer Size	9, 15 or 21
	Drop out ratio	0.25
	Learning rate	0.001
	Epoch	500
	Batch size	96
HMM	Number of HMM states	3, 5 or 7
	Number of EM Iteration	500
	Tolerance	0.001
	Initial Mean	0.10
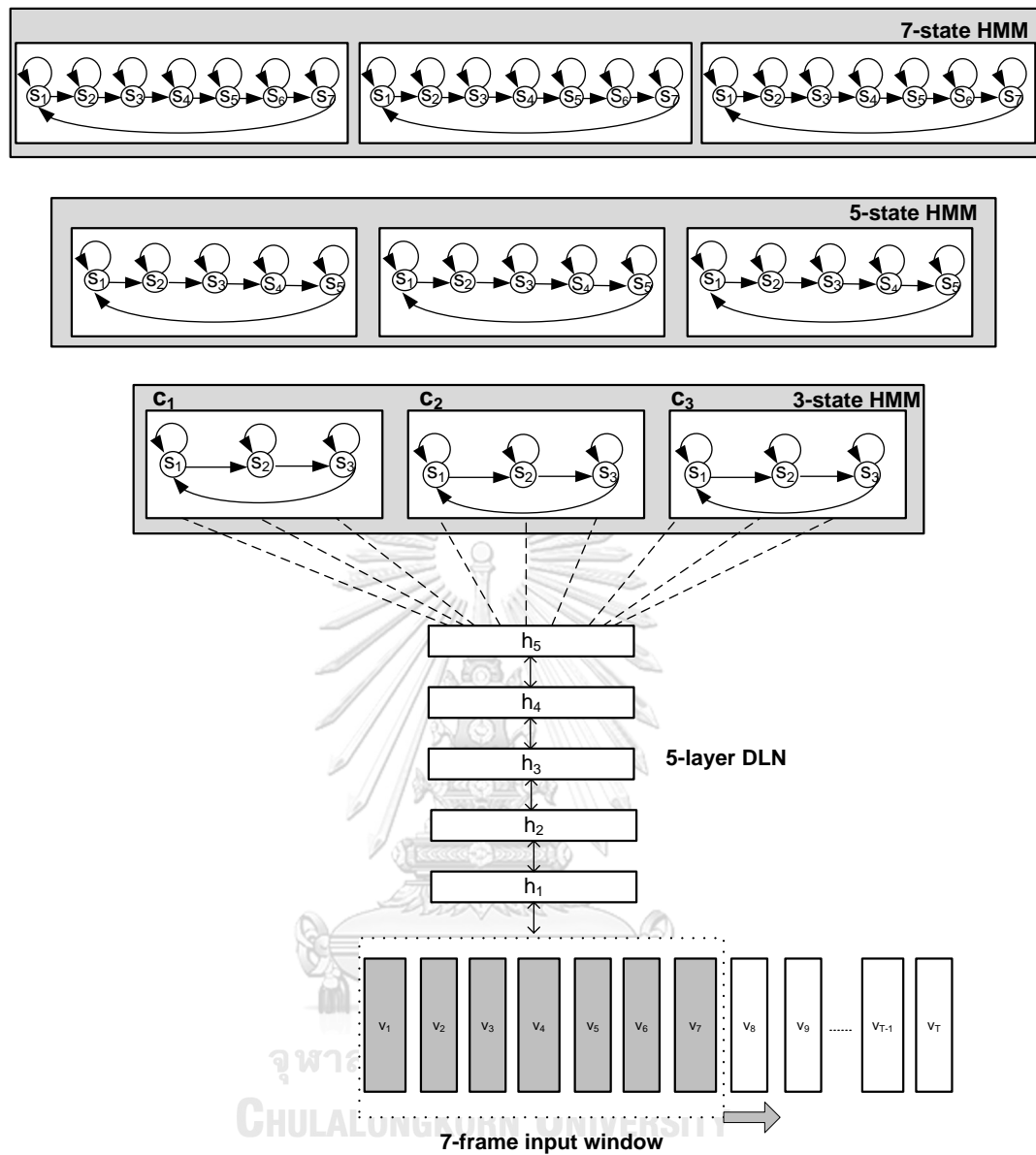
Figure 41   System Configuration for n-state Variants

The classification accuracy of valence and arousal states with n-state variants (n=3, 5 and 7) is shown in Table 5 and Table 6 respectively. The experiment performs subject-dependent classification with 10-fold cross validation in each variant of system configuration parameters. The valence and arousal classification accuracy of individual subjects are illustrated in Figure 42 and Figure 43.

Table 5   *Valence Accuracy with n-state Variants*

		#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	Mean	SD
Subj 1	n=3	58.33	63.33	54.17	62.50	66.67	64.17	69.17	66.67	65.00	66.67	63.67	4.47
	n=5	63.33	74.17	64.17	70.83	69.17	60.83	65.83	69.17	59.17	71.67	66.83	4.94
	n=7	65.83	61.67	63.33	58.33	70.83	64.17	56.67	70.83	75.83	65.83	65.33	5.91
Subj 2	n=3	64.14	62.50	59.17	70.83	66.67	63.33	70.83	61.67	60.83	72.50	65.25	4.70
	n=5	68.33	71.67	64.17	69.17	60.83	66.67	65.83	71.67	70.83	66.67	67.58	3.48
	n=7	62.50	65.00	65.83	58.33	61.67	64.17	68.33	69.17	72.50	71.67	65.92	4.52
Subj 3	n=3	59.17	64.17	62.50	63.33	64.17	68.33	57.50	58.33	62.50	64.17	62.42	3.27
	n=5	65.83	67.50	63.33	70.83	59.17	60.83	70.83	69.17	66.67	69.17	66.33	4.07
	n=7	68.33	72.50	64.17	65.83	58.33	59.17	71.67	70.83	69.17	72.50	67.25	5.26
Subj 4	n=3	58.33	61.67	54.17	57.50	55.83	59.17	60.83	55.83	57.50	53.33	57.42	2.71
	n=5	66.67	60.83	61.67	55.00	55.83	50.83	63.33	60.83	64.17	51.67	59.08	5.42
	n=7	58.33	59.17	62.50	54.17	56.67	60.83	62.50	57.50	59.17	51.67	58.25	3.46
Subj 5	n=3	60.83	64.17	59.17	60.83	53.33	56.67	63.33	68.33	64.17	55.83	60.67	4.54
	n=5	61.67	63.33	61.67	59.17	55.83	59.17	62.50	64.17	65.83	57.50	61.08	3.12
	n=7	59.17	65.83	63.33	60.83	54.17	60.00	66.67	69.17	64.17	59.17	62.25	4.43
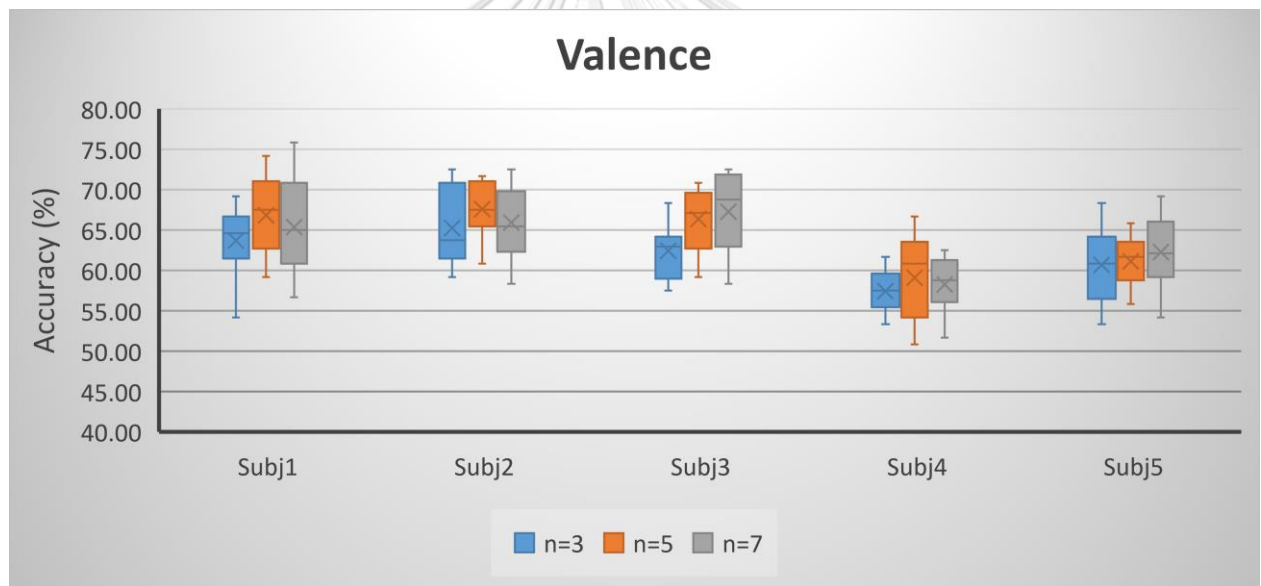
Figure 42   *Valence Accuracy of n_state Variant.*

From the experimental results, each individual subject obtained the most accurate classification for valence emotional states following: Subject1 66.58% (n=5), Subject2 67.58% (n=5), Subject3 67.25% (n=7), Subject4 59.08% (n=5) and Subject5 62.25% (n=7). The variant of n=5 provided the most accurate performance for Subject1, Subject2 and Subject4. But the variant of n=7 provided the most accurate performance for Subject3 and Subject5.

Table 6   *Arousal State Classification Accuracy with n-state Variants*

		#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	Mean	SD
	n=3	55.83	60.00	58.33	59.17	62.50	66.67	60.83	68.33	59.17	61.67	61.25	3.79
Subj 1	n=5	56.67	61.67	59.17	60.83	65.83	67.50	61.67	69.17	61.67	67.50	63.17	4.10
	n=7	59.17	63.33	59.17	58.33	71.67	72.50	65.83	70.83	68.33	66.67	65.58	5.39
	n=3	61.67	65.83	58.33	60.83	66.67	59.17	65.83	69.17	65.00	68.33	64.08	3.82
Subj 2	n=5	62.50	69.17	60.83	63.33	68.33	60.83	67.50	71.67	65.83	72.50	66.25	4.27
	n=7	60.83	70.00	62.50	64.17	64.17	59.17	66.67	65.83	66.67	69.17	64.92	3.46
	n=3	61.67	63.33	62.50	56.67	59.17	63.33	66.67	64.17	58.33	67.50	62.33	3.49
Subj 3	n=5	63.33	64.17	63.33	59.17	60.83	65.83	69.17	66.67	60.83	71.67	64.50	3.91
	n=7	65.83	63.33	64.71	60.00	60.83	68.33	72.50	68.33	63.33	73.33	66.05	4.54
	n=3	58.33	60.83	53.33	55.83	51.67	54.17	58.33	54.17	56.67	58.33	56.17	2.84
Subj 4	n=5	60.00	61.67	55.83	56.67	50.83	55.83	62.50	59.17	60.83	59.17	58.25	3.50
	n=7	62.50	60.83	55.83	59.17	53.33	57.50	63.33	63.33	64.17	63.33	60.33	3.73
	n=3	57.50	60.83	64.17	55.83	59.17	60.83	54.17	60.00	59.17	60.83	59.25	2.84
Subj 5	n=5	58.33	61.67	62.50	56.67	61.17	59.17	56.67	63.33	61.67	63.33	60.45	2.56
	n=7	60.83	63.33	65.83	58.33	60.83	62.50	58.33	64.17	62.50	64.17	62.08	2.49
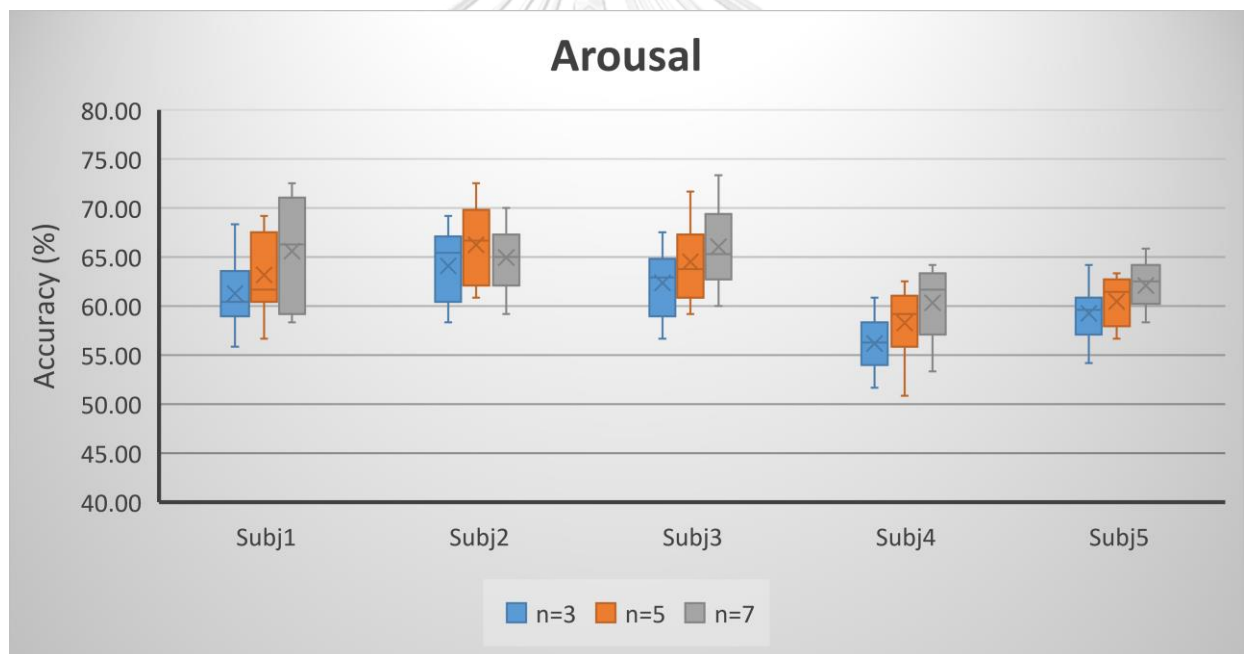
Figure 43   *Arousal Accuracy of n_state Variant.*

From the experimental results, each individual subject received the most accurate classification for arousal emotional states as following: Subject1 65.58% (n=7), Subject2 66.25% (n=5), Subject3 66.05% (n=7), Subject4 60.33% (n=7) and Subject5 62.08% (n=7). The variant of n=7 provided the most accurate performance for Subject1, Subject3, Subject4 and Subject5. But the variant of n=5 provided the most accurate performance for Subject2.

5.2.2   Variant of m-layer DLN

The number of hidden layer of the DLN (m-layer) is a key factor to enhance the DLN-HMM's accuracy performance. This section examines the variant of m-layer DLN parameters by adjusting m = 3, 5 and 7 states respectively. All of other parameters have been fixed in each variant of m-layer value. Table 7 lists all of parameters for variant of m-layer DLN. The system configuration is depicted in Figure 44.

Table 7   Configuration Parameters of m-layer DLN Variant

Parameters		Value
Input	Number of input frame window(f-frame)	7 frames
	Number of Input Features	560
DLN	Number of DLN Layers (m-layer)	3, 5 or 7 layers
	Softmax Layer Size	15
	Drop out ratio	0.25
	Learning rate	0.001
	Epoch	500
	Batch size	96
HMM	Number of HMM states	5
	Number of EM Iteration	500
	Tolerance	0.001
	Initial Mean	0.10

The classification accuracy of valence and arousal states with m-layer DLN variants (m=3, 5 and 7) is shown in Table 8 and Table 9. The experiment performs subject-dependent classification with 10-fold cross validation in each variant of system configuration parameters. The box-plot of valence and arousal classification accuracy of individual subjects are shown in Figure 45 and Figure 46 respectively.

Figure 44   System Configuration for m-layer DLN Variant

Table 8 Valence Accuracy with m-layer Variants

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	Mean	SD
m=3	59.17	68.33	61.67	68.33	63.33	57.50	59.17	68.33	60.83	65.83	63.25	4.20
m=5	63.33	74.17	64.17	70.83	69.17	60.83	65.83	69.17	59.17	71.67	66.83	4.94
m=7	61.67	70.83	60.83	67.50	66.67	58.33	60.83	65.83	60.00	69.17	64.17	4.34
m=3	64.17	68.33	63.33	70.00	58.33	64.17	62.50	66.67	68.33	65.83	65.17	3.42
m=5	68.33	71.67	64.17	69.17	60.83	66.67	65.83	71.67	70.83	66.67	67.58	3.48
m=7	67.50	66.67	63.33	68.33	59.17	66.67	64.17	68.33	68.33	66.67	65.92	2.92
m=3	62.50	65.83	61.67	68.33	60.83	55.83	68.33	67.50	64.17	66.67	64.17	4.01
m=5	65.83	67.50	63.33	70.83	59.17	60.83	70.83	69.17	66.67	69.17	66.33	4.07
m=7	65.83	66.67	61.67	70.83	59.17	58.33	69.17	68.33	65.83	68.33	65.42	4.29
m=3	63.33	59.17	58.33	54.17	51.67	50.83	60.83	58.33	61.67	50.83	56.92	4.68
m=5	66.67	60.83	61.67	55.00	55.83	50.83	63.33	60.83	64.17	51.67	59.08	5.42
m=7	64.17	58.33	60.00	53.33	52.50	51.67	62.50	61.67	62.50	52.50	57.92	4.93
m=3	57.50	60.00	60.83	55.83	53.33	60.83	58.33	58.33	60.83	54.17	58.00	2.78
m=5	61.67	63.33	61.67	59.17	55.83	59.17	62.50	64.17	65.83	57.50	61.08	3.12
m=7	63.33	64.17	60.83	57.50	56.67	58.33	58.33	61.67	64.17	55.58	60.06	3.19
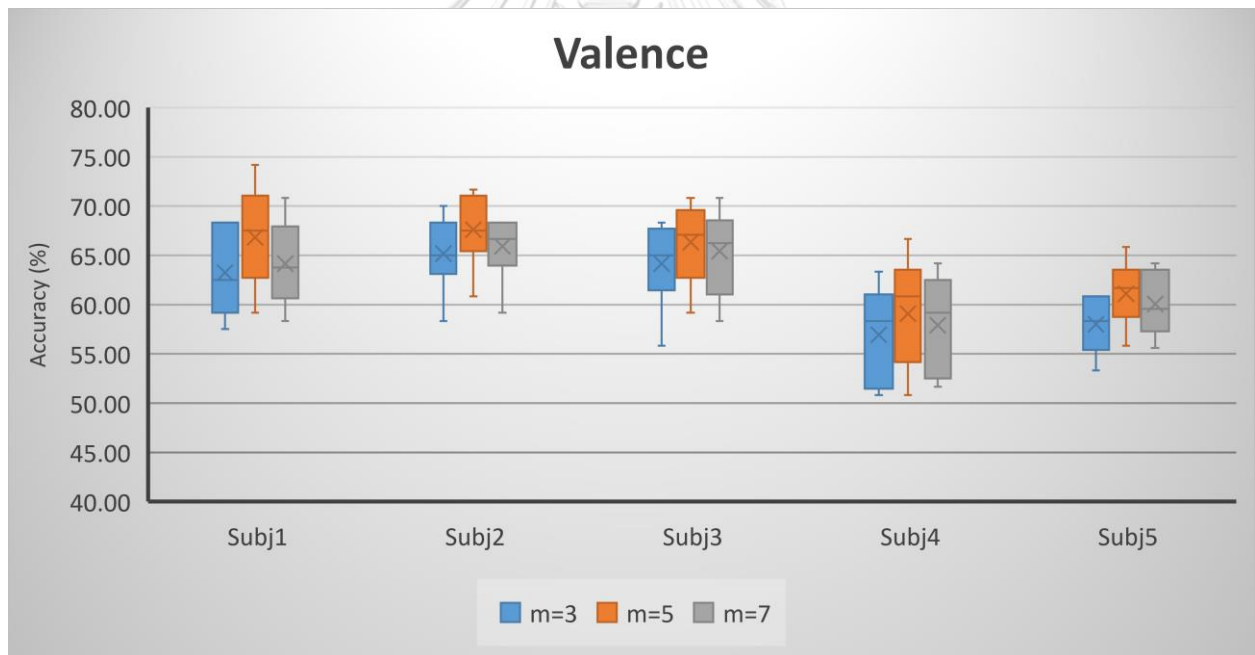
Figure 45   Valence Accuracy of m_layer Variant

From the experimental results, each individual subject obtained the most accurate classification for valence emotional states as following: Subject1 66.83% (m=5), Subject2 67.58% (m=5), Subject3 66.33% (m=5), Subject4 59.08% (m=5) and Subject5 61.08% (m=5). The variant of m=5 provided the most accurate performance for all individual subjects.

Table 9 Arousal Accuracy with m-layer Variants

|	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	Mean	SD
m=3	54.14	56.67	60.83	55.83	63.33	64.17	60.83	67.50	60.00	63.33	60.66	4.16
m=5	56.67	61.67	59.17	60.83	65.83	67.50	61.67	69.17	61.67	67.50	63.17	4.10
m=7	55.83	57.50	60.83	56.67	65.83	65.00	59.17	68.33	61.67	66.67	61.75	4.49
m=3	58.33	65.83	54.17	62.50	66.67	58.33	65.00	69.17	61.67	66.67	62.83	4.72
m=5	62.50	69.17	60.83	63.33	68.33	60.83	67.50	71.67	65.83	72.50	66.25	4.27
m=7	63.33	67.50	58.33	61.67	68.33	61.67	66.67	70.00	62.50	67.50	64.75	3.75
m=3	62.50	60.00	59.17	57.50	60.83	63.33	65.83	61.67	56.67	69.17	61.67	3.79
m=5	63.33	64.17	63.33	59.17	60.83	65.83	69.17	66.67	60.83	71.67	64.50	3.91
m=7	62.50	63.33	59.17	57.50	60.83	66.67	68.33	64.14	61.67	69.17	63.33	3.83
m=3	57.50	56.67	51.67	57.50	51.67	53.33	61.67	58.33	56.67	58.33	56.33	3.20
m=5	60.00	61.67	55.83	56.67	50.83	55.83	62.50	59.17	60.83	59.17	58.25	3.50
m=7	58.33	60.83	53.33	55.83	50.83	53.33	60.00	58.33	60.83	60.00	57.16	3.61
m=3	55.83	56.67	60.83	53.33	57.50	56.67	53.33	64.17	54.17	60.83	57.33	3.60
m=5	58.33	61.67	62.50	56.67	61.17	59.17	56.67	63.33	61.67	63.33	60.45	2.56
m=7	60.83	58.33	61.67	55.83	59.17	60.00	56.67	65.83	58.33	60.83	59.75	2.83
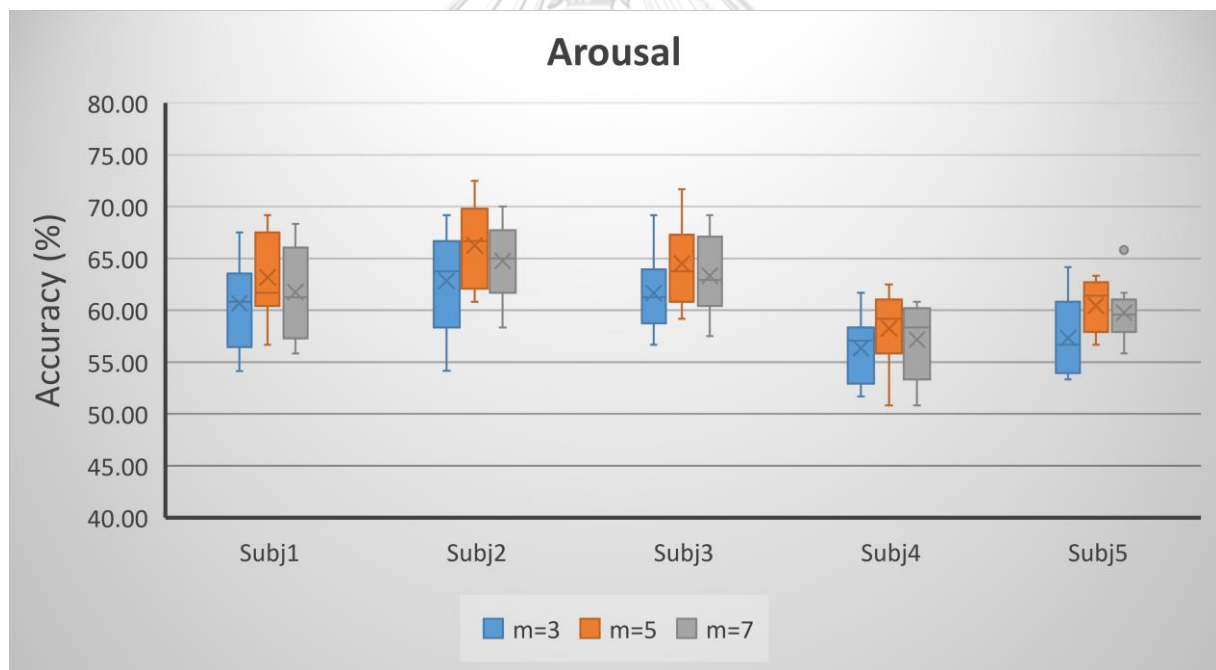
Figure 46   Arousal Accuracy of m_layer Variant

From the experimental results, each individual subject received the most accurate classification for arousal emotional states as following: Subject1 63.17% (m=5), Subject2 66.25% (m=5), Subject3 64.50% (m=5), Subject4 58.25% (m=5) and Subject5 60.45% (m=5). The variant of m=5 provided the most accurate performance for all individual subjects.

5.2.3   Variant of f-frame input window

The number of input feature frame (f-frame) is a crucial parameter to improve the DLN-HMM's accuracy performance. This section examines the variant of f-frame input window parameters by adjusting f = 3, 5 and 7 frames respectively. All of other parameters have been fixed in each variant of f-frame value. Table 10 lists all of parameters for variant of f-frame input window and Table 11 lists the number of hidden layer nodes in each f-frame variant. The system configuration is depicted in Figure 47.

Table 10   Configuration Parameters of f-frame Variant

Parameters		Value
Input	Number of input frame window(f-frame)	3, 5 or 7 frames
	Number of Input Features	240, 400 or 560
DLN	Number of DLN Layers (m-layer)	5 layers
	Softmax Layer Size	15
	Drop out ratio	0.25
	Learning rate	0.001
	Epoch	500
	Batch size	96
HMM	Number of HMM states	5
	Number of EM Iteration	500
	Tolerance	0.001
	Initial Mean	0.10

Table 11   Number of hidden layer nodes in f-frame Variant

Parameters		Value
f = 3	Input Layer size	240
	First Layer size	200
	Second Layer Size	100
	Third Layer Size	75
	Fourth Layer Size	50
	Fifth Layer Size	25
f = 5	Input Layer size	400
	First Layer size	300
	Second Layer Size	200
	Third Layer Size	100
	Fourth Layer Size	50
	Fifth Layer Size	25
f = 7	Input Layer size	560
	First Layer size	300
	Second Layer Size	200
	Third Layer Size	100
	Fourth Layer Size	50
	Fifth Layer Size	25

The classification accuracy of valence and arousal states with f-frame input window variants (f=3, 5 and 7) is shown in Table 12 and Table 13 respectively. The experiment performs subject-dependent classification with 10-fold cross validation in each variant of system configuration parameters. The box-plot of valence and arousal classification accuracy of individual subjects are shown in Figure 48 and Figure 49 respectively.
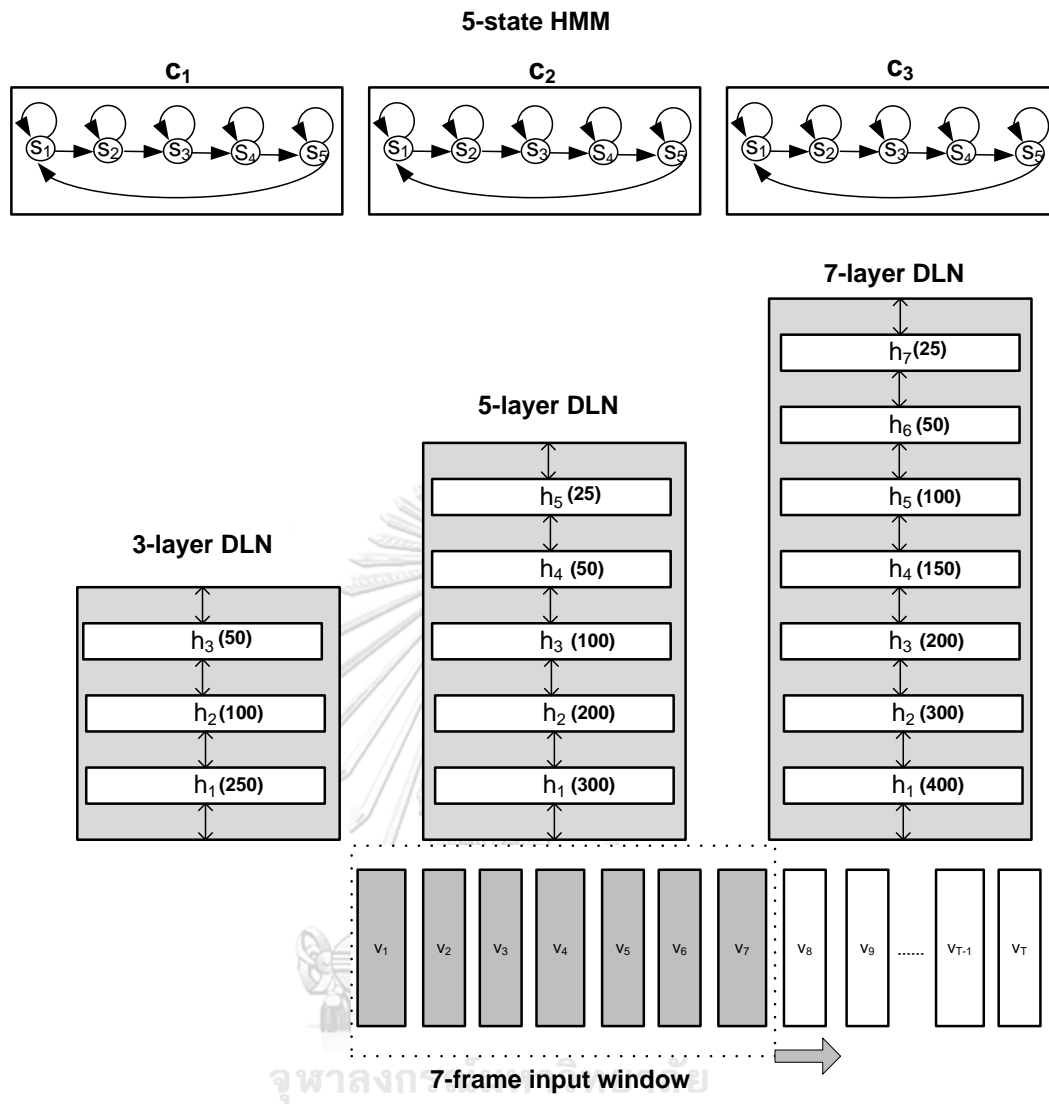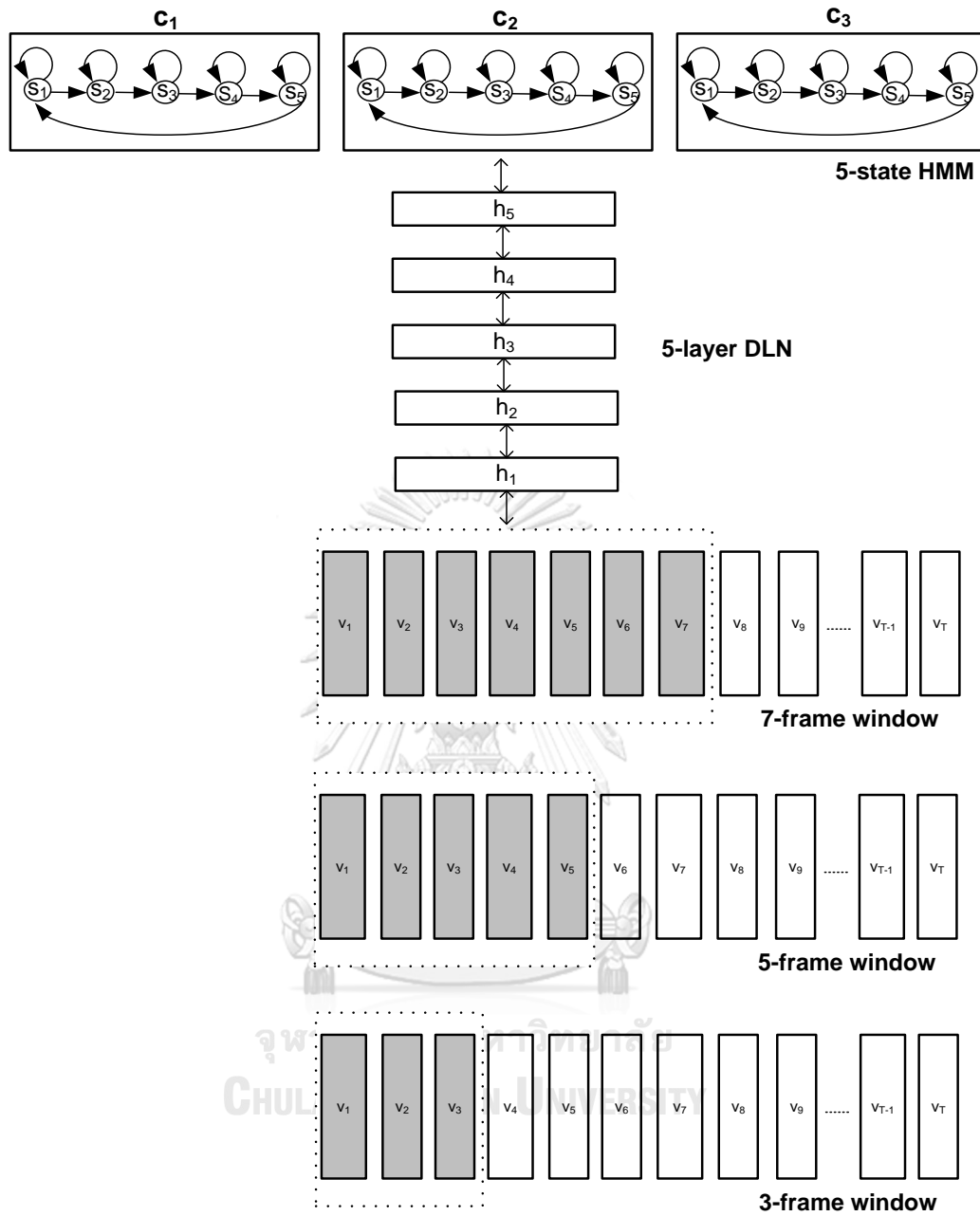
Figure 47   System Configuration for f-frame Variants

*Table 12*   *Valence Accuracy with f-frame Variants*

		#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	Mean	SD
	f=3	61.67	65.83	61.67	67.50	65.83	58.33	60.83	64.17	53.33	64.17	62.33	4.19
Subj 1	f=5	60.83	71.67	66.67	72.50	71.67	55.83	61.67	67.50	56.67	68.33	65.33	6.20
	f=7	63.33	74.17	64.17	70.83	69.17	60.83	65.83	69.17	59.17	71.67	66.83	4.94
	f=3	63.33	68.33	61.67	64.17	55.83	65.83	63.33	69.17	67.50	63.33	64.25	3.86
Subj 2	f=5	67.50	71.67	65.00	68.33	59.17	68.33	66.67	70.83	68.33	65.83	67.17	3.47
	f=7	68.33	71.67	64.17	69.17	60.83	66.67	65.83	71.67	70.83	66.67	67.58	3.48
	f=3	62.50	63.33	59.17	66.67	57.50	56.67	67.50	64.17	65.83	65.83	62.92	3.89
Subj 3	f=5	64.17	65.83	62.50	71.67	60.83	59.17	70.83	68.33	68.33	70.83	66.25	4.45
	f=7	65.83	67.50	63.33	70.83	59.17	60.83	70.83	69.17	66.67	69.17	66.33	4.07
	f=3	63.33	58.33	59.17	53.33	54.17	50.83	61.67	58.33	62.50	50.83	57.25	4.68
Subj 4	f=5	66.67	62.50	63.33	57.50	56.67	51.67	63.33	61.67	66.67	51.67	60.17	5.54
	f=7	66.67	60.83	61.67	55.00	55.83	50.83	63.33	60.83	64.17	51.67	59.08	5.42
	f=3	56.67	61.67	56.67	57.50	54.17	56.67	59.17	60.83	64.17	56.67	58.42	3.00
Subj 5	f=5	60.83	64.17	61.67	60.83	56.67	58.33	61.67	63.33	67.50	57.50	61.25	3.27
	f=7	61.67	63.33	61.67	59.17	55.83	59.17	62.50	64.17	65.83	57.50	61.08	3.12
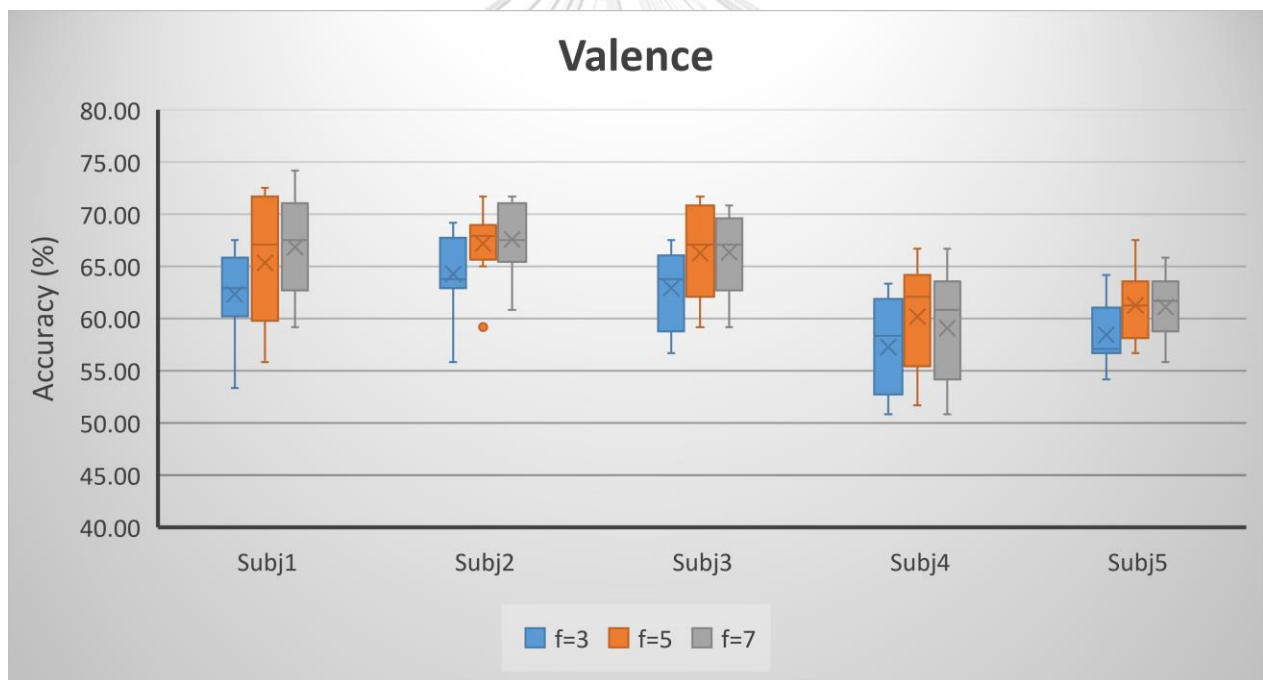
Figure 48   *Valence Accuracy of f-frame Variant*

From the experimental results, each individual subject obtained the most accurate classification for valence emotional states as following: Subject1 66.83% (f=7), Subject2 67.58% (f=7), Subject3 66.33% (f=7), Subject4 60.17% (f=5) and Subject5 61.25% (f=5). The variant of f=7 provided the most accurate performance for Subject1, Subject2 and Subject3. But the variant of f=5 provided the most accurate performance for Subject4 and Subject5.

Table 13   Arousal Accuracy with f-frame Variants

		#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	Mean	SD
	f=3	50.83	55.83	53.33	54.17	62.50	63.33	56.67	65.83	60.00	61.67	58.42	4.94
Subj 1	f=5	54.17	62.50	57.50	58.33	66.67	66.67	59.17	70.83	63.33	64.17	62.33	5.06
	f=7	56.67	61.67	59.17	60.83	65.83	67.50	61.67	69.17	61.67	67.50	63.17	4.10
	f=3	58.33	65.83	58.33	60.83	64.17	57.50	61.67	64.17	63.33	68.33	62.25	3.56
Subj 2	f=5	65.83	72.50	63.33	64.17	69.17	63.33	66.67	68.33	67.50	70.83	67.17	3.12
	f=7	62.50	69.17	60.83	63.33	68.33	60.83	67.50	71.67	65.83	72.50	66.25	4.27
	f=3	58.33	62.50	59.17	57.50	60.83	61.67	68.33	63.33	56.67	67.50	61.58	3.98
Subj 3	f=5	62.50	66.67	62.50	61.67	61.67	63.33	71.67	64.17	58.33	69.17	64.17	3.95
	f=7	63.33	64.17	63.33	59.17	60.83	65.83	69.17	66.67	60.83	71.67	64.50	3.91
	f=3	58.33	56.67	54.17	53.33	50.83	54.17	60.00	58.33	58.33	59.17	56.33	3.02
Subj 4	f=5	61.67	60.83	57.50	55.83	51.67	56.67	63.33	60.83	61.67	61.67	59.17	3.62
	f=7	60.00	61.67	55.83	56.67	50.83	55.83	62.50	59.17	60.83	59.17	58.25	3.50
	f=3	56.67	58.33	60.83	54.17	57.50	56.67	53.33	61.67	58.33	61.67	57.92	2.89
Subj 5	f=5	61.67	62.50	64.17	58.33	63.33	61.67	59.17	64.17	62.50	63.33	62.08	1.97
	f=7	58.33	61.67	62.50	56.67	61.17	59.17	56.67	63.33	61.67	63.33	60.45	2.56
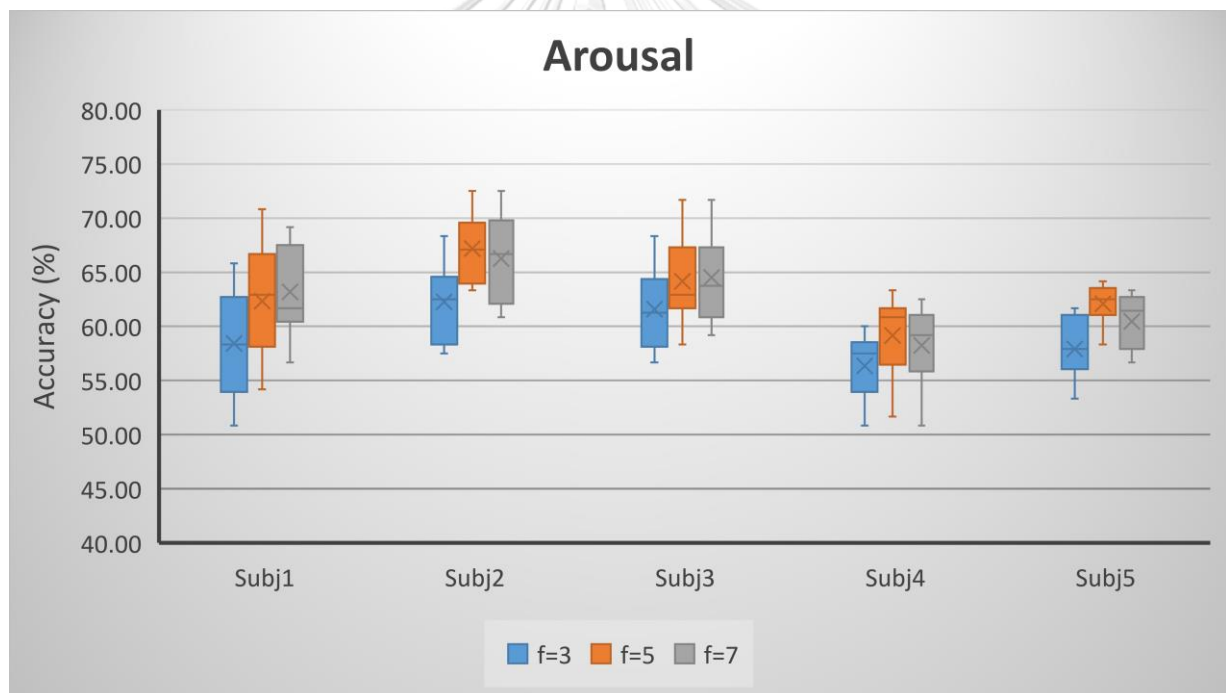
Figure 49   Arousal Accuracy of f-frame Variant

From the experimental results, each individual subject received the most accurate classification for arousal emotional states as following: Subject1 63.17% (f=7), Subject2 67.17% (f=5), Subject3 64.50% (f=7), Subject4 59.15% (f=5) and Subject5 62.08% (f=5). The variant of f=7 provided the most accurate performance for Subject1 and Subject3. But the variant of f=5 provided the most accurate performance for Subject2, Subject4 and Subject5.

5.3 Discussions

5.3.1 Variant of n_state HMM

The optimal number of hidden states in HMM models depends on the characteristics of applications or algorithms. In this section, we investigate the impacts of n_state HMM variants on our EEG-based Emotion Classification. In HMM, log likelihood of each sample data sequence can be used to evaluate its fitness under the model. Therefore, the average log likelihood of all sample data sequences in each emotional class is an indicator to measure the fitness quality of HMM models. The upper chart of Figure 50 shows the average log likelihood of each valence class (positive, neutral or negative) with n_state variants. From the charts, n=5 variant provides higher value of average log likelihood than n=3 variant. This implies that the n=5 variant provides better fitness of HMM model than the n=3 variant. Also, the average log likelihood of n=7 variant is slightly higher than those of n=5 variant. However, it is obvious that the improvement of this step-up, from n=5 to n=7, is much lower than those of another step-up, from n=3 to n=5.

Experimental results in Section 5.2.1 indicate that the number of hidden states of HMM affects the efficiency of accuracy performance. The lower chart in Figure 50 summarizes the valence classification accuracy of each individual subject with n_state variants. The increasing number of hidden states from n=3 to n=5 provides outstanding improvement of valence classification accuracy for all subjects. However, the increasing number of hidden states from n=5 to n=7 provides some mixed outcomes of the accuracy performance. Both Subject3 and Subject5 have better classification accuracy for this step-up but others have lower accuracy performance. It is possible that the overall improvement on the classification accuracy requires sufficient improvement on HMM model fitness. The quality of HMM models derived from EM algorithm must be significantly enhanced in order to provide noticeable improvement on classification accuracy.
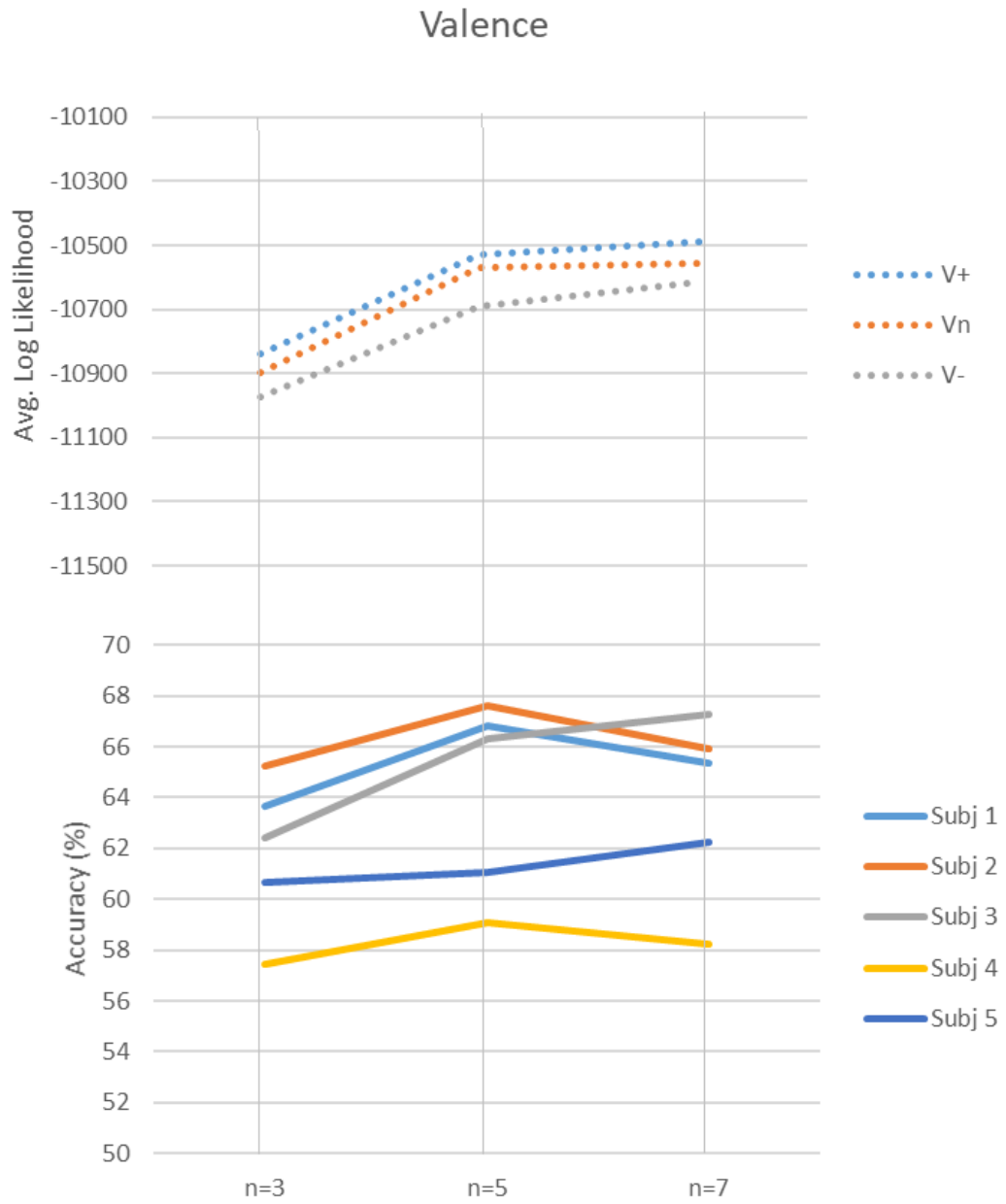
Figure 50   Average Log Likelihood vs Accuracy in Valence

In Figure 51, we found the similarity of log likelihood charts of both Valence and Arousal HMM models. The improvement of the step-up, from n=5 to n=7, is much lower than those of another step-up, from n=3 to n=5, but it is slightly better than those of Valence. In Arousal classification, most of subjects still provide better accuracy performance in the step-up, from n=5 to n=7, except Subject 1.
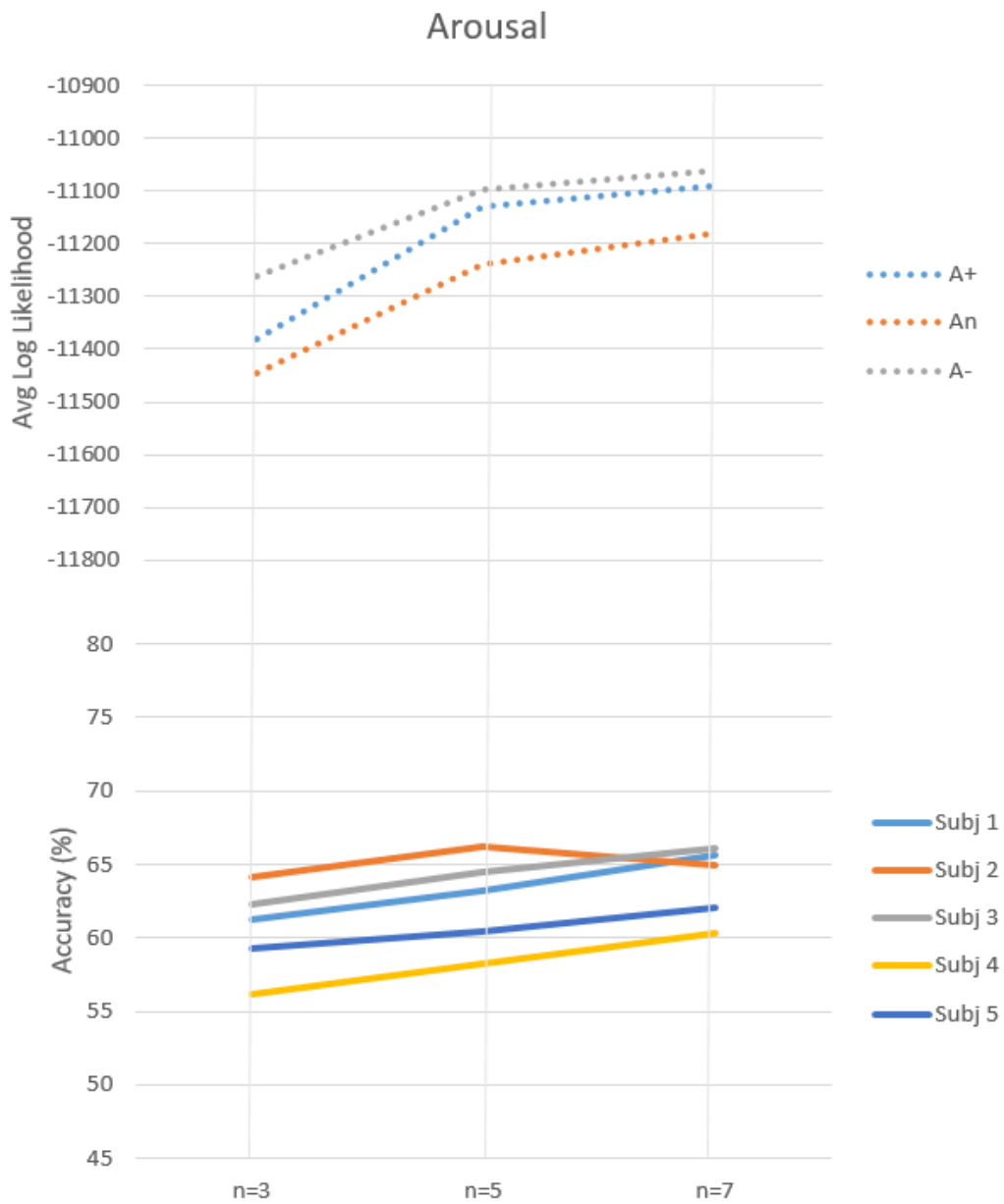
Figure 51   Average Log Likelihood vs Accuracy in Arousal

5.3.2 Variant of m_layer DLN

The DLN system configuration is crucial for estimating accuracy performance of the proposed hybrid DLN-HMM emotion classification algorithms. One of network parameters that significantly affects the accuracy performance is the number of hidden layer (m_layer). The more number of hidden layers the network has, the more complicated learning capability it provides. However, when the number of neuron nodes considerably increases, the network potentially confronts with overfitting problems.

The number of neuron nodes in each hidden layer is important for the DLN to enhance its accuracy estimation. The number of neuron nodes in each layer should be 50% to 75% of its lower layer. The number of neuron nodes in higher level should be decreased in order to force them to learn or extract more relevant features from their lower layer. Moreover, the total number of neuron nodes in the DLN should not more than twice of the number of input features. If the number of neuron nodes is overwhelming to learn its features, the network is prone to overfitting problems. In another scenario, if the number of neuron nodes is not sufficient to learn its features, the network may occur underfitting problems.

From experimental results, the emotion accuracy of individual subjects with m_layer variant for valence and arousal is summarized in Figure 52 and Figure 53 respectively. The system configuration with five hidden layers provides the most accurate classification of both valence and arousal for all subjects. The accuracy performance of m=5 variant is better than those of m=3 variant. However, when the system configuration has seven hidden layers, its accuracy performance decreases compared of m=5. But the accuracy performance of m=7 is still better than those of m=3.
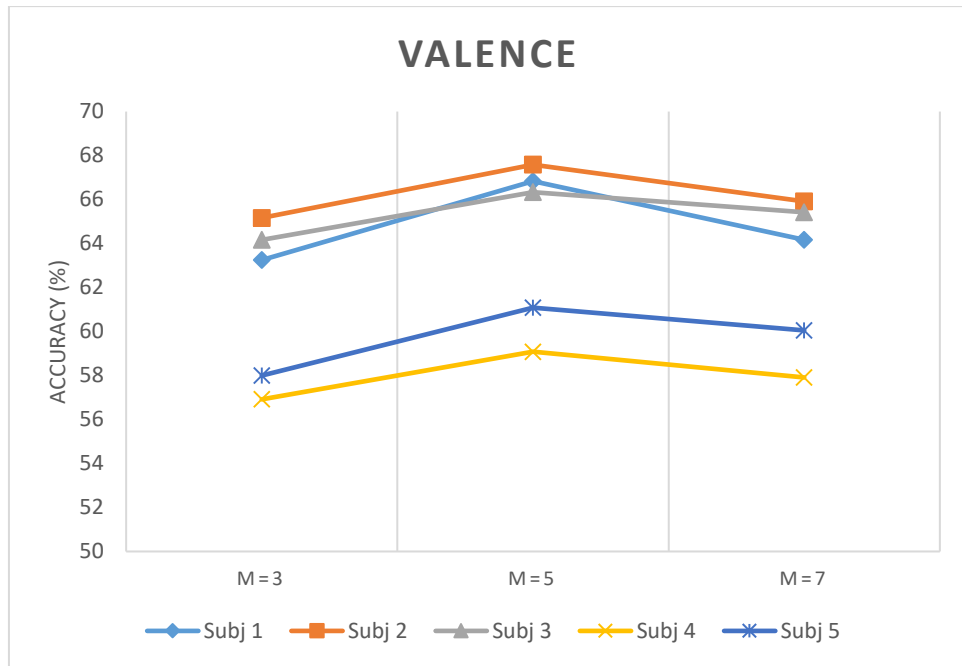
Figure 52   Valence Accuracy in m_layer Variants
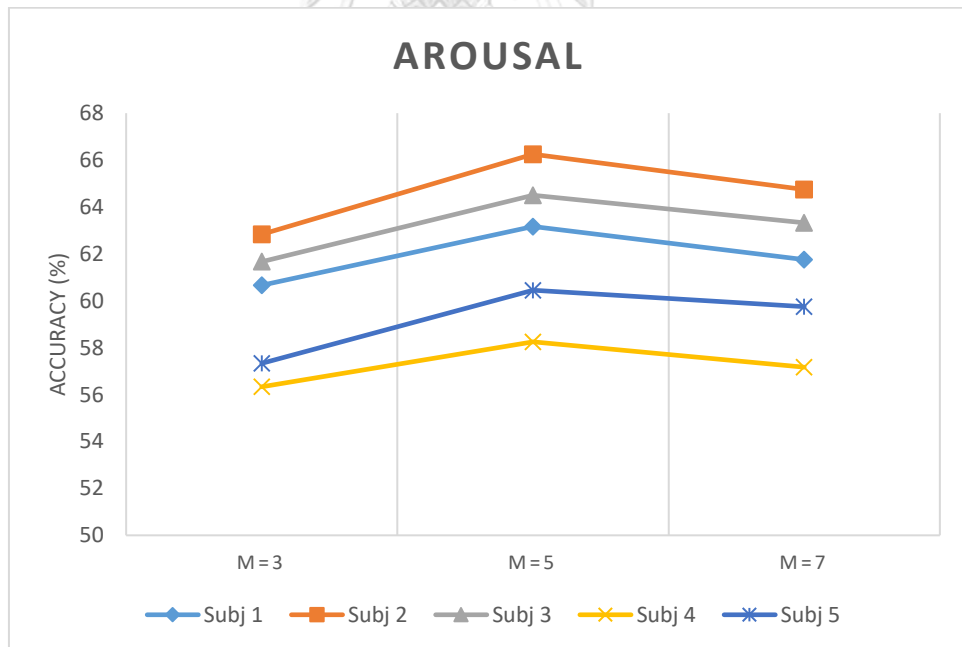
Figure 53   Arousal Accuracy in m_layer Variants

The optimal number of hidden layers in deep learning network configuration exclusively depends on its specification of application or algorithm. There is no rule of thumb to determine the optimal number of hidden layers. It relies on many parameters of the configuration such as the number of input features, the number of output nodes and the complexity of problem. Moreover, the number of neuron nodes in each hidden layers has significant impacts on its accuracy performance. There are several suggestions to perform initial approximation of these parameters. Boger et al. [70] studied the development and application of several techniques for knowledge extraction from trained ANN models. They found that the number of neuron nodes in hidden layer should be less than 2/3 of the size of input nodes. Berry et al. [71] suggested that the total number of neuron nodes in hidden layers should be less than twice of the number of input features. Blum [72] provide the guideline that the size of the hidden layer neurons should be between the input layer size and the output layer size. In this dissertation, our DLN system configuration follows these suggestions to initially approximate the number of hidden layers and the number of neuron nodes in each layer.

From experimental results, among these variants of m_layer DLN, the DLN network configuration for 5 hidden layers performs the best accuracy performance. With m=5, this variant meets all requirements in suggestions [70-72]. The number of neuron nodes is less than 2/3 of the size of input feature nodes. The total number of neuron nodes in hidden layers should be less than twice of the number of input nodes. In addition, the size of the hidden layer neurons is between the input layer size and the output layer size.

In case of m=3, this system configuration provides worst accuracy performance among the variants, even though this variant follows all of suggestions. It is possible that the number of neurons, 400 nodes (250 + 100 + 50), is not sufficient for the network to learn this complicated task. The system potentially confronts with underfitting scenario.

In m=7, this variant provides accuracy performance better than m=3 but less than m=5. This system configuration has the total number of neuron nodes is more than twice of the number of input feature nodes (560 input nodes). There are 1,225 neuron nodes (400 + 300 + 200 + 150 + 100 + 50 + 25) in this network configuration. This m=7 variant possibly occurs the overfitting scenario with the high number of hidden neuron nodes in the network.

This dissertation further investigates the impacts of the number of hidden layers on accuracy performance. Additional two experiments was setup for examining the accuracy outcomes of m=4 and m=6 network configurations, illustrated in Figure 54. The size of each hidden layers for two network configurations is depicted in the figure. These experiment setups perform the accuracy estimation of Subject1 only. The classification accuracy of valence and arousal states with m_layer variants (m=3, 4, 5, 6 and 7) is shown in Table 14 and Table 15. The valence and arousal classification accuracy are illustrated in Figure 55 and Figure 56 respectively.
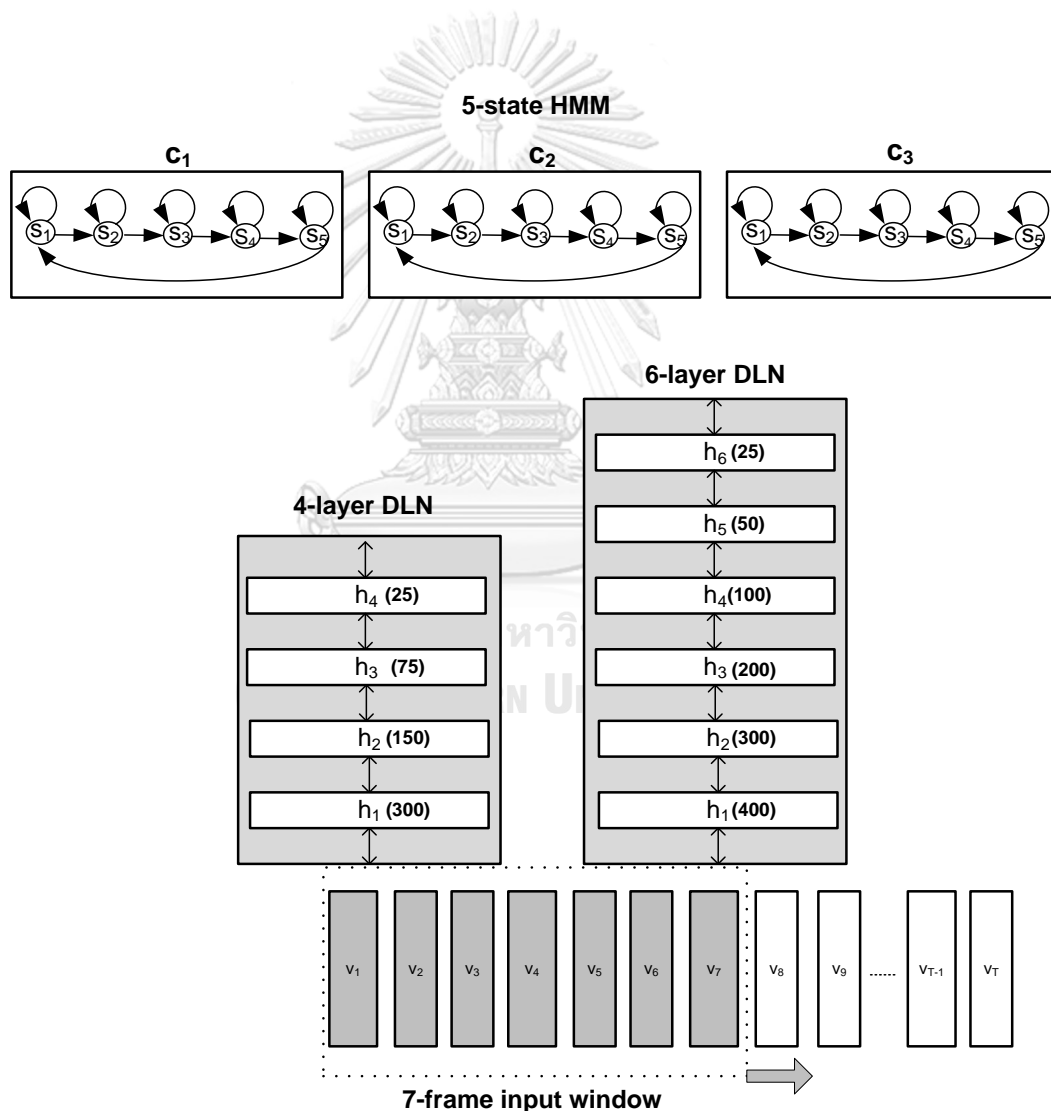
Figure 54   *Additional Experiments for m=4 and m=6 variants*

From experimental results, the optimal number of hidden layers for the proposed hybrid DLN-HMM model is five hidden layers. An additional experiment for m=4 shows that its accuracy performance is lower than those of m=5 but better than those of m=3. It is possible that this network configuration (m=4) is able to alleviate the underfitting problem. There are more numbers of hidden neuron nodes in the network. The total number of neuron nodes is 550 nodes (300 + 150 + 75 + 25 nodes). They can collaborate to learn complicated features.

In case of m=6, its accuracy performance is lower than those of m=5 but better than those of m=7. The total number of neuron nodes for m=6 configuration is 1075 nodes (400 + 300 + 200 + 100 + 50 + 25 nodes). The total number of neuron nodes is still less than twice of the number of input feature nodes. However, the accuracy is less than those of m=5 configuration. It is possible that the m=6 starts to encounter with the overfitting problem, even though the total number of neuron nodes is still less than twice of the number of input feature nodes. The optimal number of hidden layers depends on the application specific.

Table 14   *Valence Accuracy Performance of m_layer variants (Subject1)*

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	Mean	SD
m=3	59.17	68.33	61.67	68.33	63.33	57.50	59.17	68.33	60.83	65.83	63.25	4.20
m=4	61.67	72.50	61.67	71.67	65.85	59.17	62.50	67.50	56.67	66.67	64.59	5.17
m=5	63.33	74.17	64.17	70.83	69.17	60.83	65.83	69.17	59.17	71.67	66.83	4.94
m=6	62.50	72.50	63.33	68.33	66.67	61.67	64.17	67.50	60.83	69.17	65.67	3.76
m=7	61.67	70.83	60.83	67.50	66.67	58.33	60.83	65.83	60.00	69.17	64.17	4.34
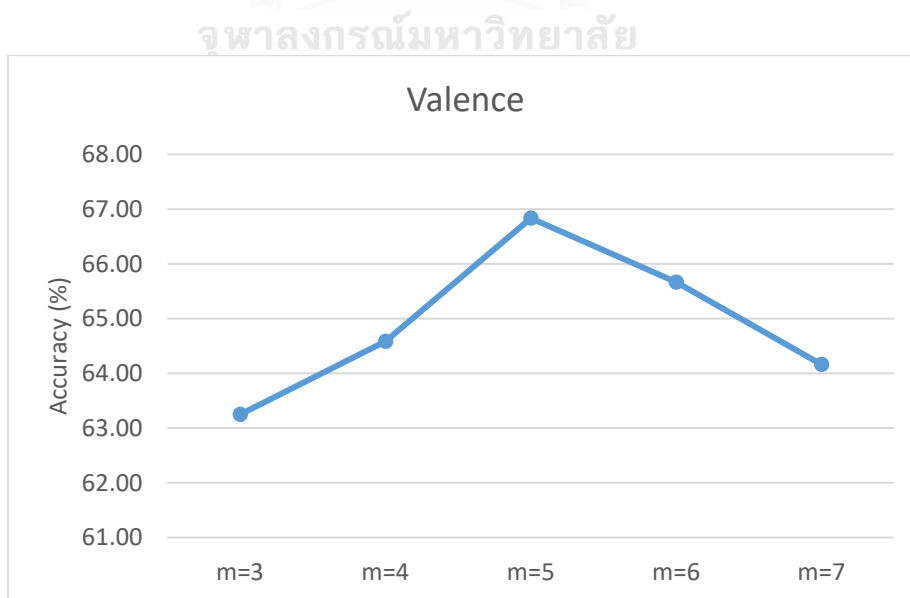
Figure 55   *Valence Accuracy of m_layer Variant (Subject1)*

Table 15   *Arousal Accuracy of m_layer variants (Subject1)*

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	Mean	SD
m=3	54.14	56.67	60.83	55.83	63.33	64.17	60.83	67.50	60.00	63.33	60.66	4.16
m=4	54.14	60.00	59.17	59.17	64.17	66.67	60.83	66.67	62.50	66.67	62.00	4.13
m=5	56.67	61.67	59.17	60.83	65.83	67.50	61.67	69.17	61.67	67.50	63.17	4.10
m=6	57.50	59.17	57.50	60.83	64.17	67.50	61.67	69.17	60.00	66.67	62.42	4.22
m=7	55.83	57.50	60.83	56.67	65.83	65.00	59.17	68.33	61.67	66.67	61.75	4.49
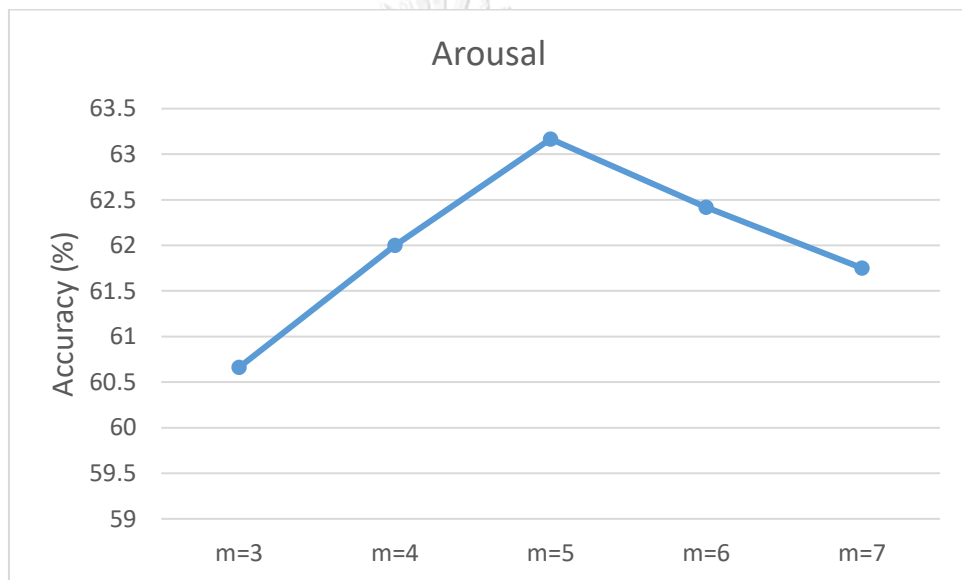
Figure 56   *Arousal Accuracy of m_layer Variant (Subject1)*

5.3.3 Variant of f_frame Input Window Size.

The purpose of this section is to examine the impacts of the number of consecutive input feature frames on the hybrid DLN-HMM's accuracy performance. There are three experimental setups of f_frame input window sizes (f_frame = 3, f_frame = 5 and f_frame = 7), described in Section 5.2.3. In these experiment setup, the number of hidden HMM states is fixed to n=5 and the number of hidden layer is fixed to m=5. Therefore, the numbers of neuron nodes in each hidden layer in the DLN have been adjusted to be congruous with their input feature size – 240, 400 and 560, as shown in Table 11.

The experimental results for valence emotion classification of each individual subject are summarized in Figure 57. From the experimental results, there is significant improvement on accuracy performance when the number of input frames is increased from f=3 to f=5 for all subjects. However, the setup from f=5 to f=7 provides mixed outcomes of the accuracy. Only Subject1 provided better accuracy from 65.33% to 66.83%. Only Subject4 provided worse accuracy from 60.17% to 59.08%. Other subjects provided almost same accuracy.
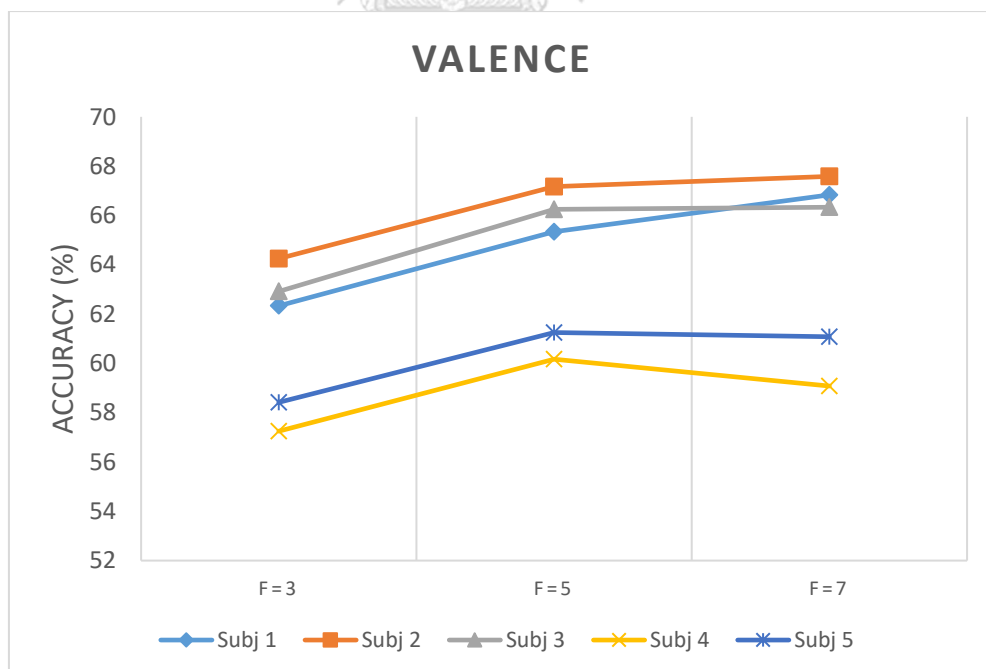
Figure 57   Valence Accuracy in f_frame Variants

Figure 58 illustrates the experimental results for arousal emotion classification of each individual subject. Like valence, there is significant improvement on accuracy performance when the number of input frames is increased from f=3 to f=5 for all subjects. However, the setup from f=5 to f=7 provides mixed outcomes of the accuracy. Subject1 and Subject3 provided better accuracy. Other subjects provided worst accuracy.
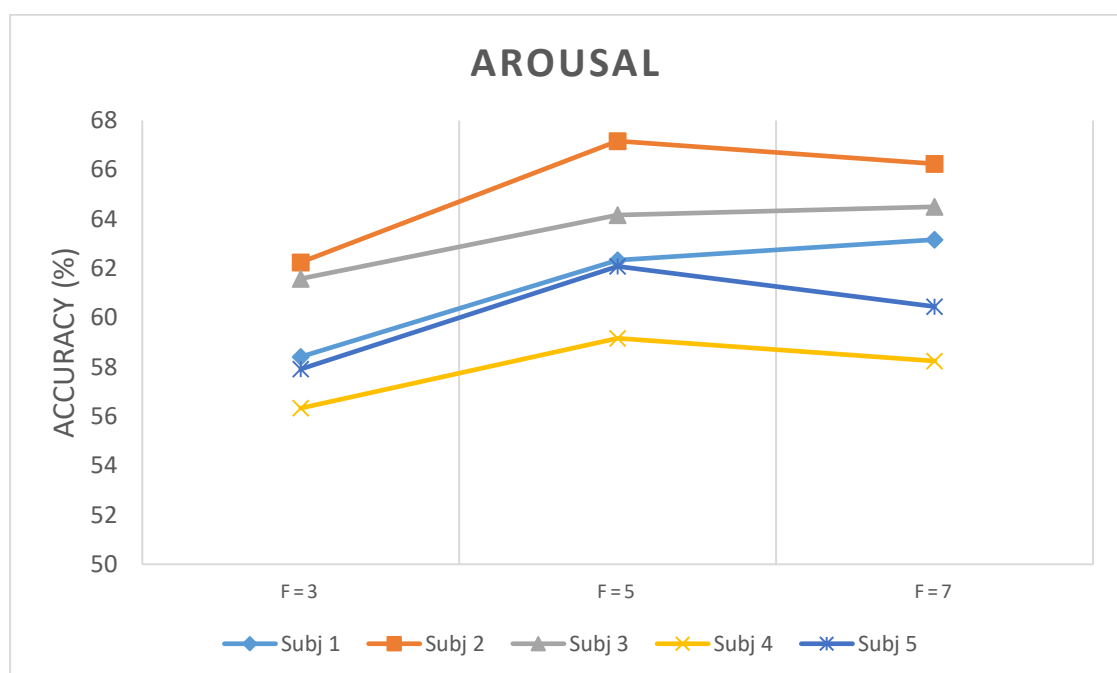
Figure 58   Arousal Accuracy in f_frame Variants

The concept of multi-frame input windowing allows the DLN to learn input features of consecutive input frames simultaneously. This overlapping of input feature frames is similar to perform 1-dimensional Convolution Neural Network (1D CNN) computation in the input layer. The CNN well performs for identifying or recognizing simple patterns in low-level features and then these features will be employed to formulate more complex patterns in high-level layers. The object recognition in 2D CNN can learn the object's edges at low-level layers and then it combines all of recognized edges of the object at high-level layers.

The 1D CNN is considerably effective to extract relevant features from shorter segments of the whole sequence in time-series. This methodology performs extracting relevant features within a particular input window. The window of input features will travel from the beginning through the end of the input feature sequence of a sample data. By increasing from f=3 to f=5, both valence and arousal classifications provide better accuracy performance for all subjects. It seems that the algorithm is able to learn features better with enlarged input windows. However, by increasing from f=5 to f=7, the outcomes of accuracy performance were mixed. It is possible that the optimal number of input frames depends on the temporal dynamics of brain wave activities of individual subjects.

Chapter 6

Static vs Dynamic Approach Comparison

6.1 Overview of Comparison

This research study intensively examines the potential of deep learning networks to perform hierarchical feature learning of EEG-based emotion classification system. We implement the emotion classification system in two approaches – Static and Dynamic- to classify the cognitive states of emotion using deep learning networks.

The proposed system with static approach is implemented with a stack of fully connected hidden layers and softmax layers on top of the network. The system performs emotion classification by estimating valence and arousal states from softmax outputs. In the Static approach, all labels of feature input sequences are the same in one data sample, derived from one emotion stimulus. For example, all input feature sequences are extracted from one emotion stimulus by watching a video and then their labels will be labelled for one of emotion classes.

This dissertation further develops another version of our EEG-based emotion classification system with Dynamic approach. The system exploits the concept of temporal neural dynamics of emotion to enhance the accuracy performance of the proposed system. The procedure of emotional activation in brain involves with three temporal dynamic features of an emotional response to a single event – Emotion onset, Emotion duration and Emotion Resurgence, described in Section 2.8. The proposed algorithm utilizes Hidden Markov Models (HMMs) to learn meaningful transitions of high-level features deriving from the DLNs. In Dynamic approach, the labels of feature input sequences are the optimal sequences of mapping states generated by GMM-HMM models.

The main purpose of this chapter is to investigate the accuracy performance improvement on the Dynamic approach over the Static approach. We perform statistical analysis to examine the difference of these two approaches by paired t-test computation. We are not able to compare these two approaches by using the outcomes of static approach in Section 4.2 because this DLN version was implemented with our own proprietary implementation with Matlab which is not be implemented with Keras-Tensorflow platform. Therefore, another system needs to be implemented with Keras to compare the efficiency between these two approaches.

6.2 System Implementation

This version of EEG-based emotion classification system is constituted of a stack of hidden layers with a softmax layer on top. The DLN networks for valence and arousal classification are separated for their own hierarchical feature learning capability. The network of valence has three softmax nodes for each valence level (positive, neutral and negative). Also, the network of arousal has three softmax nodes for each arousal level (active, neutral and passive), illustrated in Figure 59.
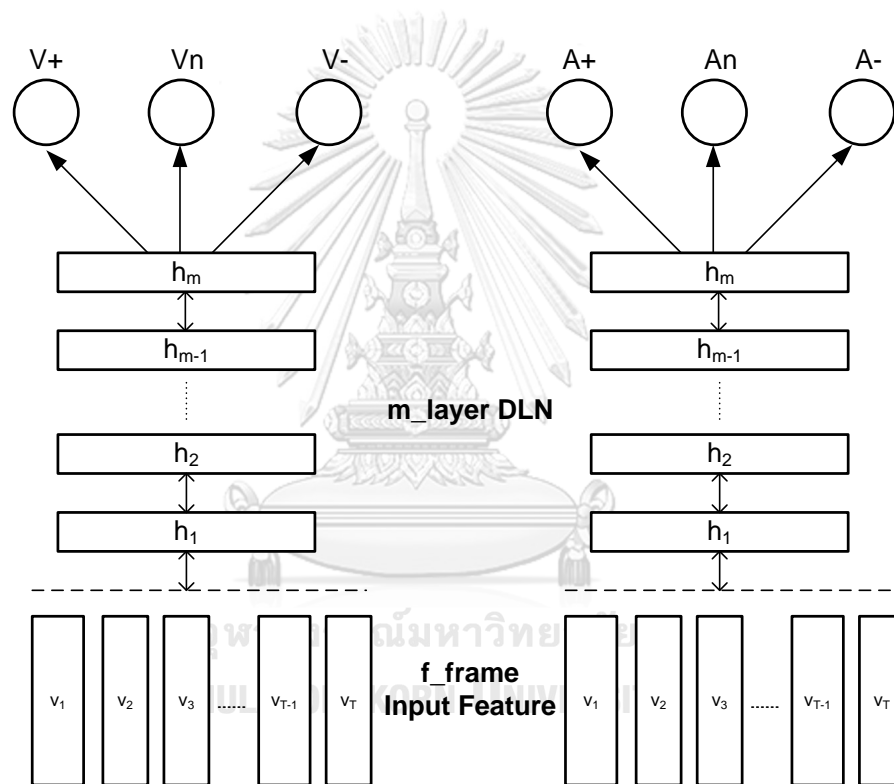
Figure 59   Static DLN System Architecture with Keras-Tensorflow.

6.2.1 Variants of m_layer DLN

This section measures the accuracy performance of each m-layer DLN variant by adjusting m = 3, 5 and 7 states respectively. All of other parameters have been fixed in each variant of m-layer value and they are same values of Dynamic approach in Section 5.2.2, except the number of softmax nodes. Table 16 lists all of parameters for variant of m-layer DLN. The system configuration is depicted in Figure 60.

Table 16   Parameters of m-layer DLN Variant without HMMs

Parameters		Value
Input	Number of input frame window(f-frame)	7 frames
	Number of Input Features	560
DLN	Number of DLN Layers (m-layer)	3, 5 or 7 layers
	Softmax Layer Size	3
	Drop out ratio	0.25
	Learning rate	0.001
	Epoch	500
	Batch size	96

จุฬาลงกรณ์มหาวิทยาลัย

CHULALONGKORN UNIVERSITY

Figure 60   *System Configuration for m-layer DLN Variant without HMMs*

6.2.2 Variant of f_frame input window

This section measures the accuracy of each f-frame parameter by adjusting f = 3, 5 and 7 frames respectively. All of other parameters have been fixed in each variant of f-frame value and they are same values of Dynamic approach in Section 5.2.3, except the number of softmax nodes. Table 17 lists all of parameters for variant of m-layer DLN and

Table 11 lists the number of hidden layer nodes in each f-frame variant. The system configuration is depicted in Figure 61.

Table 17   Parameters of f-frame Variant without HMMs

Parameters		Value
Input	Number of input frame window(f-frame)	3, 5 or 7 frames
	Number of Input Features	240, 400 or 560
DLN	Number of DLN Layers (m-layer)	5 layers
	Softmax Layer Size	3
	Drop out ratio	0.25
	Learning rate	0.001
	Epoch	500
	Batch size	96

Figure 61   System Configuration for f-frame Variants without HMMs

6.3 Experimental Results

This experiment examines how the system's configuration without HMMs affects the emotion classification accuracy. We adjust several configuration parameters, which are the number of DLN hidden layers (m-layer) and the number of input feature frames (f-frame window). The experiment performs subject-dependent classification with 10-fold cross validation in each variant of system configuration parameters and these accuracy values are presented in "DLN" column in Table 18 and Table 19.

Table 18   Accuracy of m_layer Variants with paired t-test.

Variant	Subject	Valence			Arousal		
		DLN	DLN+HMM	*p*	DLN	DLN+HMM	*p*
m = 3	Subj1	58.86%	63.25%	*0.000173*	57.18%	60.66%	*0.000988*
	Subj2	60.75%	65.17%		57.82%	62.83%	
	Subj3	60.24%	64.17%		56.59%	61.67%	
	Subj4	54.36%	56.92%		53.23%	56.33%	
	Subj5	54.12%	58.00%		54.92%	57.33%	
m = 5	Subj1	61.73%	66.83%	*0.001171*	58.54%	63.17%	*0.000871*
	Subj2	62.88%	67.58%		61.41%	66.25%	
	Subj3	61.02%	66.33%		60.16%	64.50%	
	Subj4	56.87%	59.08%		56.11%	58.25%	
	Subj5	57.85%	61.08%		57.32%	60.45%	
m = 7	Subj1	60.21%	64.17%	*0.001031*	57.63%	61.75%	*0.000559*
	Subj2	63.38%	65.92%		60.82%	64.75%	
	Subj3	61.67%	65.42%		58.25%	63.33%	
	Subj4	54.26%	57.92%		54.78%	57.16%	
	Subj5	58.39%	60.06%		56.48%	59.75%	

Table 18 shows the experimental results of the static DLN (without HMMs) for each individual subject. In addition, the experimental results of the dynamic DLN+HMM in Section 5.2.2 are shown in column "DLN+HMM". The purpose of this chapter is to compare the accuracy performance derived from two approaches with paired t-test. We calculate p values of each m_layer variants (m=3, m=5 and m=7). From the table, the p values of all m_layer variants is less than 0.05 (p < 0.05). Therefore, the dynamic approach (DLN+ HMM) provides significant improvement compared to the static approach (DLN).

Table 19   Accuracy of f_frame Variant with paired t-test.

Variant	Subject	Valence			Arousal		
		DLN	DLN+HMM	*p*	DLN	DLN+HMM	*p*
f = 3	Subj1	56.15%	62.33%	*0.002492*	55.84%	58.42%	*0.002785*
	Subj2	60.39%	64.25%		58.54%	62.25%	
	Subj3	58.14%	62.92%		57.42%	61.58%	
	Subj4	54.06%	57.25%		55.27%	56.33%	
	Subj5	56.45%	58.42%		54.86%	57.92%	
f = 5	Subj1	60.87%	65.33%	*0.001830*	57.76%	62.33%	*0.004598*
	Subj2	62.76%	67.17%		62.83%	67.17%	
	Subj3	60.68%	66.25%		60.51%	64.17%	
	Subj4	58.21%	60.17%		57.15%	59.17%	
	Subj5	58.31%	61.25%		60.92%	62.08%	
f = 7	Subj1	61.73%	66.83%	*0.001171*	58.54%	63.17%	*0.000871*
	Subj2	62.88%	67.58%		61.41%	66.25%	
	Subj3	61.02%	66.33%		60.16%	64.50%	
	Subj4	56.87%	59.08%		56.11%	58.25%	
	Subj5	57.85%	61.08%		57.32%	60.45%	

Table 19 shows the experimental results of the static and the dynamic approaches for each individual subject with f_frame variants (f=3, f=5 and f=7). From the table, the p values of all variants is less than 0.05 (p < 0.05). Therefore, the dynamic approach (DLN+HMM) provides significant improvement compared to the static approach (DLN) for both valence and arousal emotion states.

6.4 Discussions

6.4.1 Accuracy Enhancement

This chapter emphasizes on the accuracy performance comparison of two approaches (Static vs Dynamic). In static approach, the system configuration is implemented with a stack of fully connected hidden layers and softmax layers on top of the network. This approach performs emotion classification by estimating valence and arousal states from softmax outputs. In dynamic approach, we apply the concept of temporal neural dynamics of emotion to enhance the accuracy performance of the proposed system. This approach utilizes HMM to learn meaningful transitions of high-level features deriving from the DLNs. From experimental results, the dynamic DLN+HMM approach provides better accuracy performance than the static DLN approach in both valence and arousal classifications.

From statistical analysis with paired t-test, the p values, computed by comparing all subjects in each variant, indicate that the dynamic DLN+HMM configuration is able to enhance the accuracy performance of all m_layer and f_frame variants, compared to the static DLN configuration. It is obvious that the concept of temporal neural dynamic of emotion processing is important to enhance the accuracy performance of the proposed system. HMMs are responsible for learning all relevant state transitions of high-level features in the DLN.

6.4.2 HMM vs RNN

There are two prominent methodologies to implement time-sequence state learning algorithms – Hidden Markov Model (HMM) and Recurrent Neural Network (RNN). HMM is a class of probabilistic graphical model that allow us to predict a sequence of unknown variables from a set of observed variables. The HMM can be presented as a state sequence with observations made at a sequence of time steps and it is used to predict the best sequence of hidden states. The HMM is constructed with an inference model based on assumptions of Markov process. The Markov process assumption is that the current state solely depends on its previous state. In other words, prediction of the future state does not require any other historical information except its present state information.

An alternative methodology used to learn temporal state sequences is Recurrent Neural Network (RNN). The RNN is a class of artificial neural network and their connections between neural nodes are presented as a directed graph along a temporal sequence. This configuration provides its capability to learn temporal dynamic behavior of its input sequence. RNN is a special kind of neural networks. Unlike feed forward neural network, RNNs implement internal states as their memory to process sequences of inputs and they use their memory to learn sequential process from historical information, not only from current state information.

Generally, Recurrent Neural Network offers more complex sequential state learning than Hidden Markov Model because the former has capability of learning sequential state process from information in past but the latter employs the information of its previous state only. However, the RNN has a major problem about vanishing gradient [73, 74]. In theory, RNN is able to perform temporal state learning of a long sequence very effectively. In fact, when RNN operates a very long sequence, substantial changes in the network's parameters will induce a very small change in the network's outputs of the last element in the sequence. Therefore, the network is not able to perform sequential state learning effectively. In this research study, the proposed algorithm needs to perform sequential state transition learning of temporal neural dynamics in very long state sequence (96 states). For this reason, the HMM should be more suitable to implement temporal state transition learning algorithm for this research study.

Another drawback of RNN is lack of efficiency to perform sequential state learning with small number of datasets. RNN provides the capability of learning sequential state process from long-term historic information. The complexity computation of such long-term dependency is considerably difficult to learn from small number of data. The algorithm requires sufficient amount of training datasets to learn relevant sequential state transitions [75]. The sufficient number of datasets for training the RNN depends on the application specifics. After considerations of both issues on our proposed implementation – very long state sequence and small number of datasets, we decided to implement the sequential state learning with HMM, instead of RNN.

6.4.3 Number of Datasets.

One of requirements for deep neural networks to effectively perform hierarchical feature learning algorithm is the sufficient amount of training data. The DLN's features at high-level can be learned from compositions of its features at low-level with greedy layer-by-layer unsupervised pre-training. This unsupervised pre-training allows the DLN to discover unknown feature coherences of input features and the number of the DLN's training data is crucial to provide efficient accuracy performance.

Our data collection protocol is described in Section 5.1.3. The dataset is constituted of raw EEG signals derived from picture-based emotion stimulus sessions. Each subject has total number of 200 data samples of each emotion – Happy, Relax, Sad and Afraid. Also, there are 400 data samples of Neutral emotion. For valence classification, the experiment used its positive valence data samples from Happy and Relax emotion, its negative valence data samples from Sad and Afraid, as well as its neutral valence data samples from Neutral emotion. For arousal classification, we use the active arousal data from Happy and Afraid, and passive arousal data from Relax and Sad, as well as neutral arousal from Neutral emotion. Each data sample consists of 96 sequences of 16ch x 5 EEG power bands so the dimension of input features of the DLN is 96x80 for each sample. The total number of data samples for training and testing data in the DLN is 1200 data samples and each sample has the dimension of 96x80 input features.

The purpose of this section is to examine whether the number of data samples of our experiment dataset is sufficient amount or not. An additional experiment is established to further investigate the impact of the number of data samples in our proposed system. This additional experiment requires only one subject to perform more EEG data collection sessions. The number of data samples of each emotion is increased from 200 samples to 400 sample for Happy, Relax, Sad and Afraid emotion. Also those of Neutral emotion is increased from 400 to 800 samples. Therefore, there are double size of training and testing data for the DLN in this additional experiment and all system configuration parameters are same values used in Section 5.2. This experiment performs emotion classification with 10-fold cross validation in each variant of system configuration parameters.

The summary of accuracy performance for valence and arousal emotion classification is shown in Table 20. The values of accuracy performance in column "N=1200" are derived from the experiment results in Section 5.2. These accuracy values are compared with the accuracy values from this additional experiment by increasing the number of data samples from N=1200 to N=2400 for both valence and arousal emotion classifications. From p values in the Table 20, the effect of increasing the number of data samples does not significantly enhance the system's accuracy performance. However, a majority of variants provide the accuracy performance of N=2400 experiment slightly better than those of N=1200 experiment. A minority of variants – Arousal classification with n=5, m=5, f=7 and Valence classification with f=5 -- provides the accuracy performance of N=2400 experiment slightly worse than those of N=1200 experiment.

The detail of accuracy performance in each trial on 10-fold cross validation for all variants in shown in Table 21 to Table 26 in the Appendix. Table 21 and Table 22 show the values of accuracy performance in each trial with n_state parameter variants for valence and arousal classification respectively. For m_layer variants, Table 23 and Table 24 show the values of accuracy performance for valence and arousal classification respectively. The values of valence and arousal accuracy performance in each trial with f_frame parameter variants are shown in Table 25 and Table 26 respectively.

Table 20   Accuracy of Data Sample Size Comparison

Variant	Valence				Arousal			
	N=1200	N=2400	*p*	*pᵥ*	N=1200	N=2400	*p*	*pᵥ*
n=3	63.67%	63.83%	*0.355381*		61.25%	61.96%	*0.103413*	
n=5	66.83%	67.08%	*0.268136*	*0.130410*	63.17%	63.04%	*0.391648*	*0.141673*
n=7	65.33%	65.79%	*0.204874*		65.58%	65.96%	*0.258660*	
m=3	63.25%	63.54%	*0.259464*		60.66%	61.21%	*0.154582*	
m=5	66.83%	67.08%	*0.268136*	*0.115480*	63.17%	63.04%	*0.391648*	*0.279828*
m=7	64.17%	64.46%	*0.240939*		61.75%	61.79%	*0.459767*	
f=3	62.33%	62.50%	*0.360344*		58.42%	59.00%	*0.063772*	
f=5	65.33%	65.08%	*0.287349*	*0.409727*	62.33%	62.75%	*0.200866*	*0.212615*
f=7	66.83%	67.08%	*0.268136*		63.17%	63.04%	*0.391648*	

p_v: p value of all variants in a particular parameter.

Chapter 7
EEG Power Correlations with Emotions

This dissertation explores the temporal neural dynamics of emotion processes by investigating relevant state transitions of high-level features learnt by DLNs. From the experimental results in Section 5.2, the hybrid DLL-HMM method has potential to learn EEG spectral power features and classify the cognitive states of emotion effectively. The purpose of this chapter is to discover the common patterns of EEG power correlations of particular emotional states for individual subjects. It is crucial to examine the dynamic changes of EEG spectral powers in each band. Therefore, time-course analysis of EEG power correlation is important to obtain understandings of the temporal neural dynamics of emotion processes.

Many researches [76-84] presented intensive statistical analyses on EEG power correlations with emotions and revealed significant positive/negative correlations at electrode locations on the scalp. However, the outcomes of these researches provide a variety of conclusions. For examples, some works found a significant correlation at particular electrode locations but other works did not indicate such a correlation. Consequently, this dissertation will investigate the temporal changes of EEG spectral power with emotional states on our dataset. The difference of EEG spectral power of each band at a particular location compared with its baseline will be examined throughout each data sample.

In this section, we employ the same feature extraction protocol described in Section 5.1.4 to calculate EEG spectral powers. However, these spectral powers were computed with non-overlap 1-second window size, and then these spectral powers were subtracted with its power baseline derived from the average power of neutral state before the stimulus. Subsequently, we computed the average of these EEG powers differed from its baseline on all data samples of its emotion class, and then they were normalized into [-1,1] range to illustrate its relative power synchronization/ desynchronization among electrode locations.

The EEG power correlations with valence and arousal emotions of five bands (Theta, Low Alpha, High Alpha, Low Beta and High Beta) of Subject1 is illustrated in Figure 62. For valence, we found the power synchronization (increasing power) of Theta band at temporal-central lobe (T7-C3) of left hemisphere as well as the power synchronization of both Alpha bands at left frontal lobe (F3) and occipital lobe (Oz). These outcomes are congruous with Koelstra [58]. Moreover, there occurred power

synchronization of Low and High Beta bands at F3 and F3-C3 respectively, similar to Aftanas [76]. Also, there were the power desynchronization (decreasing power) of Low Beta band at T8, similar to Kroupi [82].

For arousal, there existed Theta's power desynchronization at central-temporal lobe in right hemisphere (C4-T8) and Theta's power synchronization at occipital lobe (Oz), similar to outcomes of Koelstra [58] and Aftanas [76] respectively. The power synchronization of both Low and High Alpha bands can be found at F4 and C3 but its opposite relation can be found at Pz and Oz. Moreover, there was minor power desynchronization of Low Beta at F4-C4 area and the power desynchronization of High Beta was found at FP1.

Figure 62   EEG Power Correlations of Subject1

The EEG power correlations with valence and arousal emotions of Subject2 are illustrated in Figure 63. For valence, we found significant power desynchronizations of Theta band at F3 and F4, as well as its synchronization at PO7 and PO8. Moreover, the increasing power of both Low and High Alpha bands were found around Pz area. These outcomes are consistent with Hu [80]. The power desynchronization of both Low and High Beta bands were also found around P3-Cz area, similar to Koelstra [58]. For arousal, there existed Theta's power synchronization at the middle of parietal lobe (Pz), consistent with Hu [80]. In both Alpha bands, we also found their decreasing power at Fp1-F3 and C4-T8, and their increasing power at C3 and Fp2, congruous with Kroupi [82].
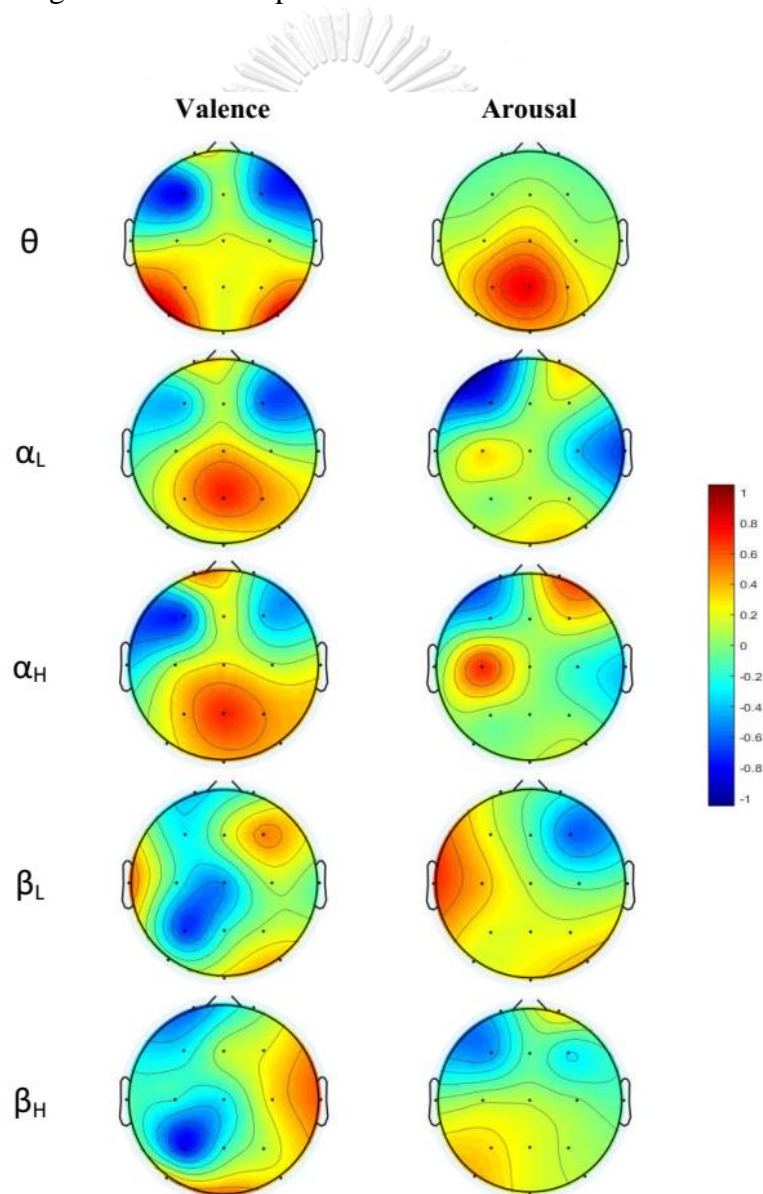
Figure 63   EEG Power Correlations of Subject2

The EEG power correlations with valence and arousal emotions of Subject3 are illustrated in Figure 64. For valence, we found the power synchronization of Theta band at left temporal-central lobe (T7-C3) and occipital lobe (Oz-PO8). Also, there occurred power synchronizations of both Alpha bands at left frontal lobe (F3) and occipital lobe (Oz). These outcomes are consistent with Koelstra [58]. There were power synchronizations of Low and High Beta bands at C3 and PO8, and minor power desynchronization at T8, similar to Kroupi [82]. For arousal, we found Theta's power synchronization at P3 and its power desynchronization at T8. The increasing power at C3-P3 area and decreasing power at Cz area can be found for both Low and High Alpha. These outcomes of Theta and Alpha are congruous with Koelstra [58]. Also, there were the decreasing power of Low Beta at F4 and High Beta at Fp1.
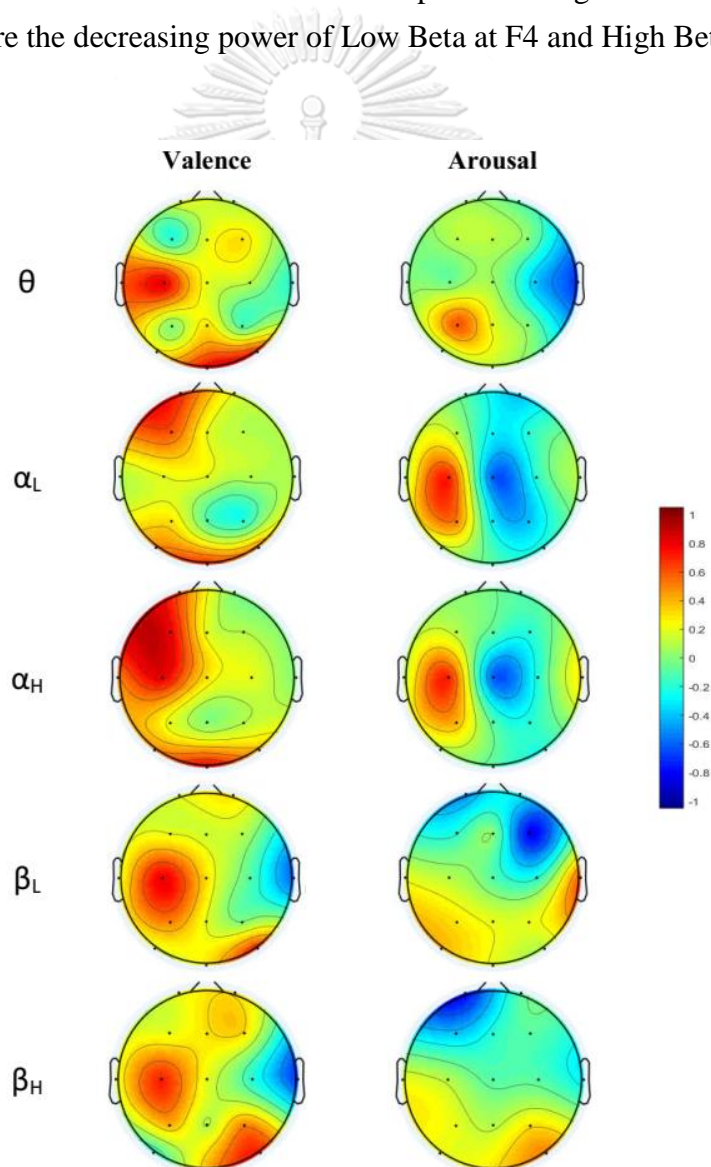
Figure 64   EEG Power Correlations of Subject3

The EEG power correlations with valence and arousal emotions of Subject4 are illustrated in Figure 65. For valence, there were a couple of power desynchronizations of Theta band at the frontal lobe in both hemispheres as well as its power synchronization at occipital area. We found the power desynchronization at pre-frontal lobe and the synchronization at occipital lobe in both Low and High Alpha bands. For arousal, we found minor desynchronization at T8 and synchronization at Pz-Oz area. In High Alpha, we found a power synchronization at C3 and minor desynchronization at Oz. We also found a minor desynchronization at F4-C4 area. There were no significant change in High Beta band for both valence and arousal emotions.
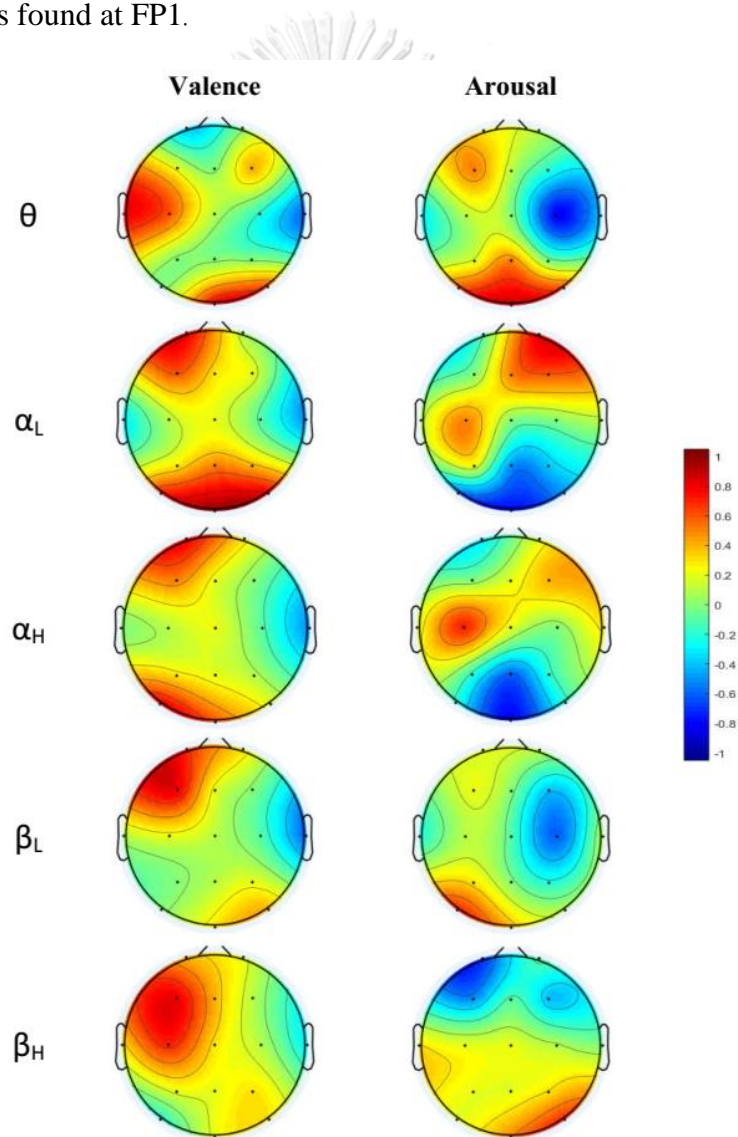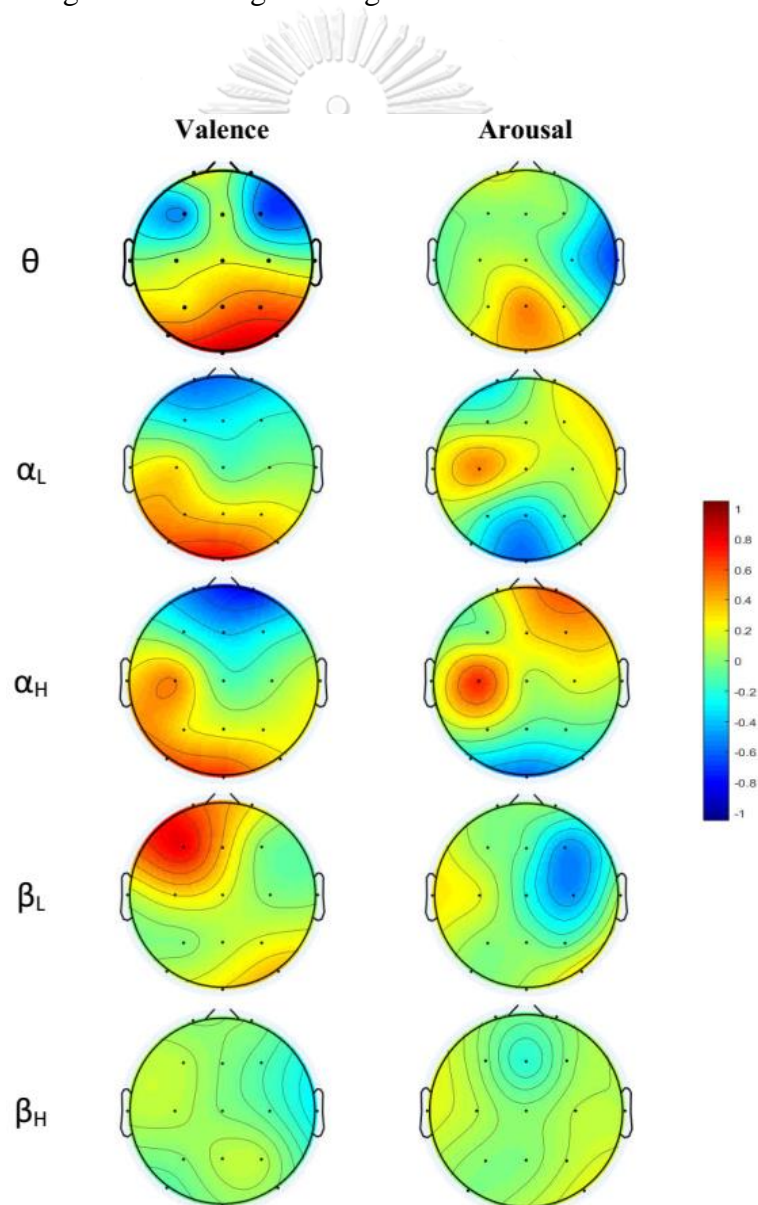

Figure 65   EEG Power Correlations of Subject4

The EEG power correlations with valence and arousal emotions of Subject5 are illustrated in Figure 66. For valence, we found the power synchronization of Theta band at pre-frontal lobe (Fp1) and occipital lobe (PO7) of left hemisphere and its desynchronization at T8, similar to Koelstra [58]. There existed power synchronization at Fp1 and Oz and desynchronization at F4-C4 area of both Low and High Alpha bands. For arousal, we found a couple of Theta power synchronization at Fp3 and Oz. The power of Low Alpha was decreased at Fp3 and T8 with minor increased power at Fp2. In Low Beta, its power desynchronization existed at Fp1 and its synchronization at Oz. There were no significant power change found in High Beta band.
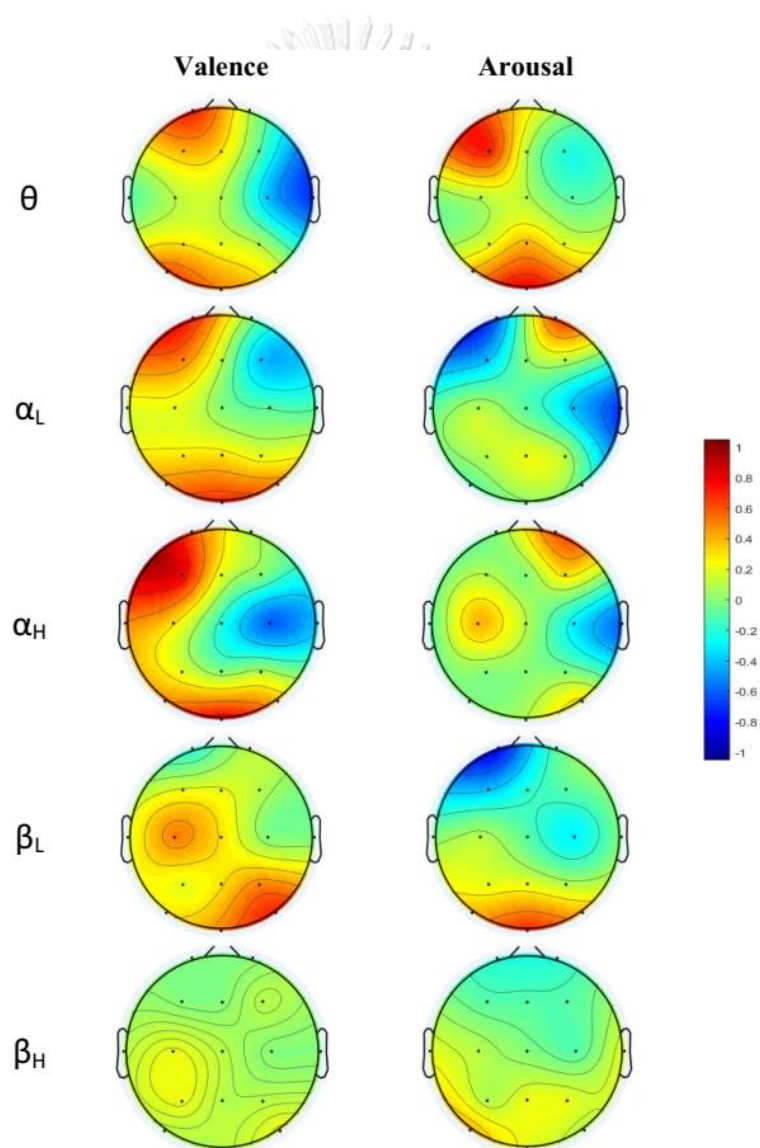
Figure 66   EEG Power Correlations of Subject5

For our dataset, the outcomes of EEG power correlation analyses of each individual subject reveal a variety of EEG spectral power changes of valence and arousal emotion compared with its baseline neutral emotion. The subjects provided their own unique patterns of EEG power correlation. Therefore, it is difficult to obtain good accuracy performance of EEG-based emotion classification with subject-independent protocol. We decided to evaluate the classification accuracy of our proposed systems with subject-dependent protocol in all experimental setups, except the experiments in Section 4.2.

The purpose of this chapter to obtain understandings about temporal neural dynamics of emotion processes. This research further performs time-course analyses of EEG power correlations in order to discover common patterns of EEG power changes of particular emotional states for individual subjects. We examines the average of EEG spectral powers with various time intervals of all data samples in each emotion class. However, the outcomes of the time-course EEG power correlations are inconsistent to determine common patterns of EEG power dynamic changes. As we know, the characteristics of EEG signals have very high variation and non-stationary. Therefore, we cannot discover such a common pattern with simple computation. It is possible that there exists a number of relevant patterns within one emotion but DLNs and HMMs are capable of learning and classifying these patterns.

One example of temporal dynamics of EEG power correlations with valence emotion is illustrated in Figure 67. These sequences of topological plots were derived from the computation of relative EEG spectral power within one data sample of Subject1. The average powers of each band and location were computed with 5-second interval windows. From topological plots, during the first 5 seconds, there occurred power synchronizations of Theta and Alpha bands at the occipital lobe and those of Low and High Beta bands at the frontal lobe in the left hemisphere. Moreover, there occurred power desynchronizations of Theta at Fp1-F3 area and those of High Alpha and Low Beta at T8 location. Subsequently, during 6-10 second period, Theta's power had been increased at T7-C3 area. There were power synchronization of both Low and High Alpha bands at Fp1 and PO7-Oz. For Low Beta, its power synchronization at F3 had been decreased from the first 5-second period. However, its power synchronization at PO8 had been increased from the first 5-second period. There occurred power synchronization of High Beta at C3 location. After 10-second, there were no significant change of EEG power correlations for all EEG bands.
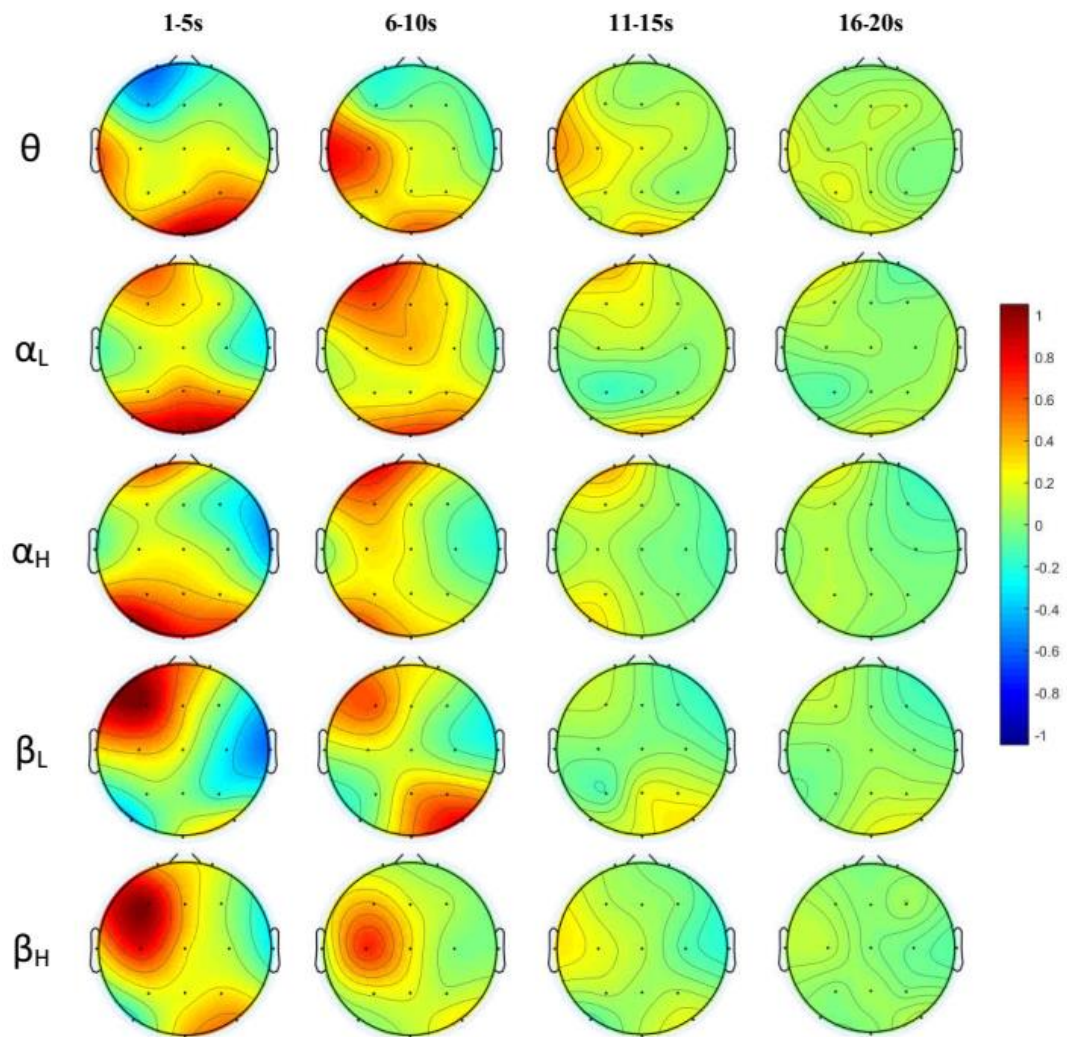
Figure 67   Time-course EEG Power Correlation of Subject 1 (Valence)

Time-course EEG power correlations with arousal emotion of Subject1 is depicted in Figure 68. During first 5 seconds, the power of Theta band had been decreased from its baseline at right-central lobe but it had been increased at occipital lobe. For Low and High Alpha bands, there existed power synchronizations at C3 and F3-T8 area but there were power desynchronization at occipital lobe. Moreover, the power of Low Beta had been decreased at F4-C4 area. For High Beta, there existed power desynchronization at frontal lobe and synchronization at PO7. During 6-10 second period, the power synchronization and desynchronization was attenuated from the first 5 second period in most of EEG bands. After 10-second, there were no major change of EEG power correlations for all EEG bands.

Figure 68   Time-course EEG Power Correlation of Subject 1 (Arousal)

Time-course EEG power correlations with valence emotion of Subject3 is depicted in Figure 69. During first 5 seconds, Theta's power had been increased from its baseline at left central lobe and right occipital lobe. For both Alpha bands, there was power synchronization at pre-frontal lobe at left hemisphere. For both Beta bands, there existed power synchronization at C3 and PO8 and desynchronization at T8. During 6-10 second, the power synchronization of Theta, Low Alpha and High Alpha bands still existed with small attenuation but the power of Theta had been slightly decreased at F3 and P4 and the power of Low Alpha had been decreased at Pz-P4 area. For Beta bands, its power synchronization and desynchronization was slightly attenuated from the first period. After 10-second, there were no major change of EEG power correlations for all EEG bands.
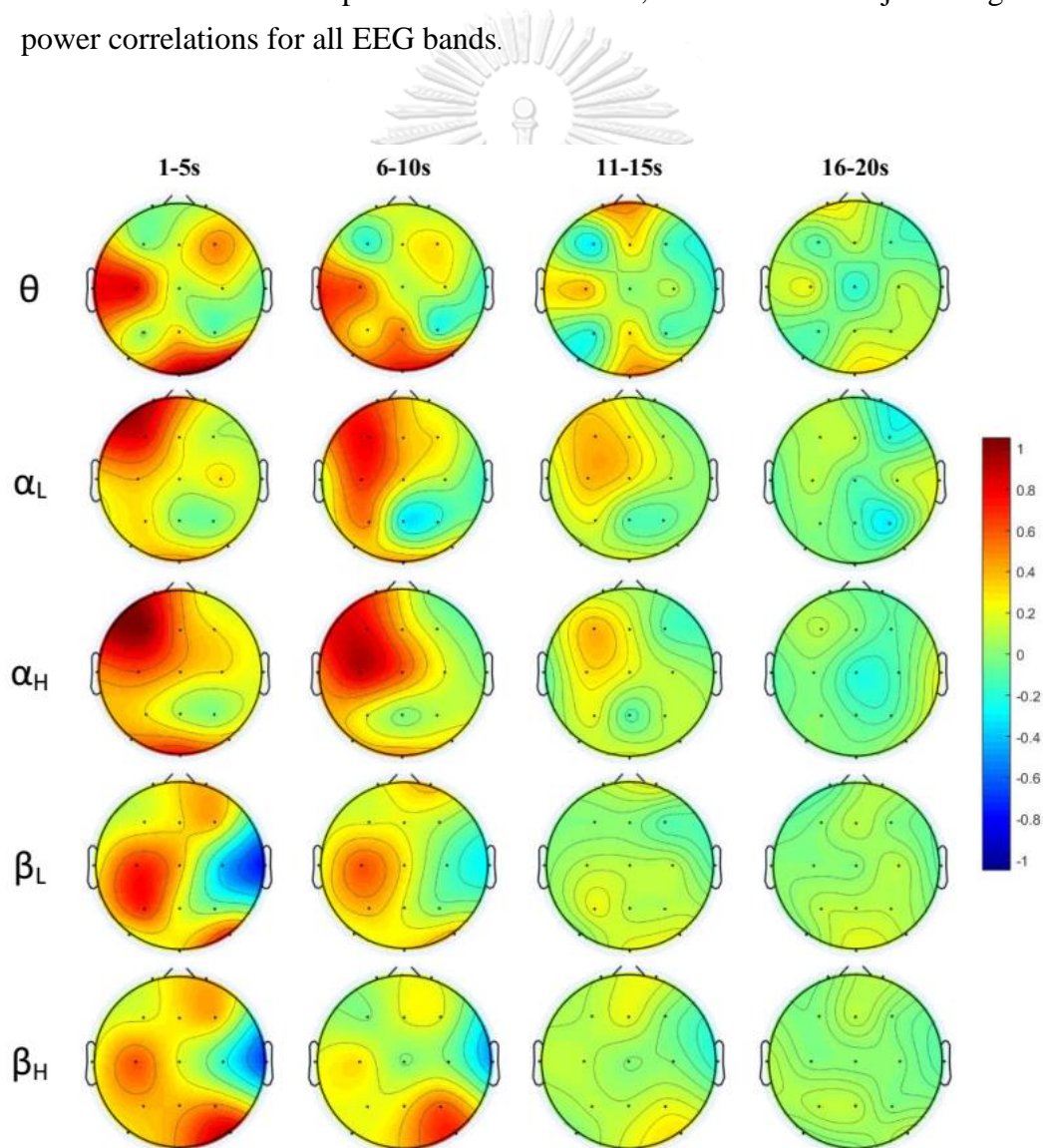
Figure 69   Time-course EEG Power Correlation of Subject 3 (Valence)

Time-course EEG power correlations with arousal emotion of Subject3 is illustrated in Figure 70. During first 5 seconds, Theta's power had been increased from its baseline at left parietal lobe and right temporal lobe. For both Alpha bands, there was power synchronization at frontal and central lobes at left hemisphere and significant desynchronization at C4 area. Moreover, there was power desynchronization of Low Beta at right frontal lobe (F4) and minor desynchronization at Fp1. For High Beta, there existed power desynchronization at Fp1 and minor synchronization at PO7 and PO8. During 6-10 second period, the power synchronization and desynchronization was attenuated from the first 5 second period in most of EEG bands. After 10-second, there were no major change of EEG power correlations for all EEG bands.

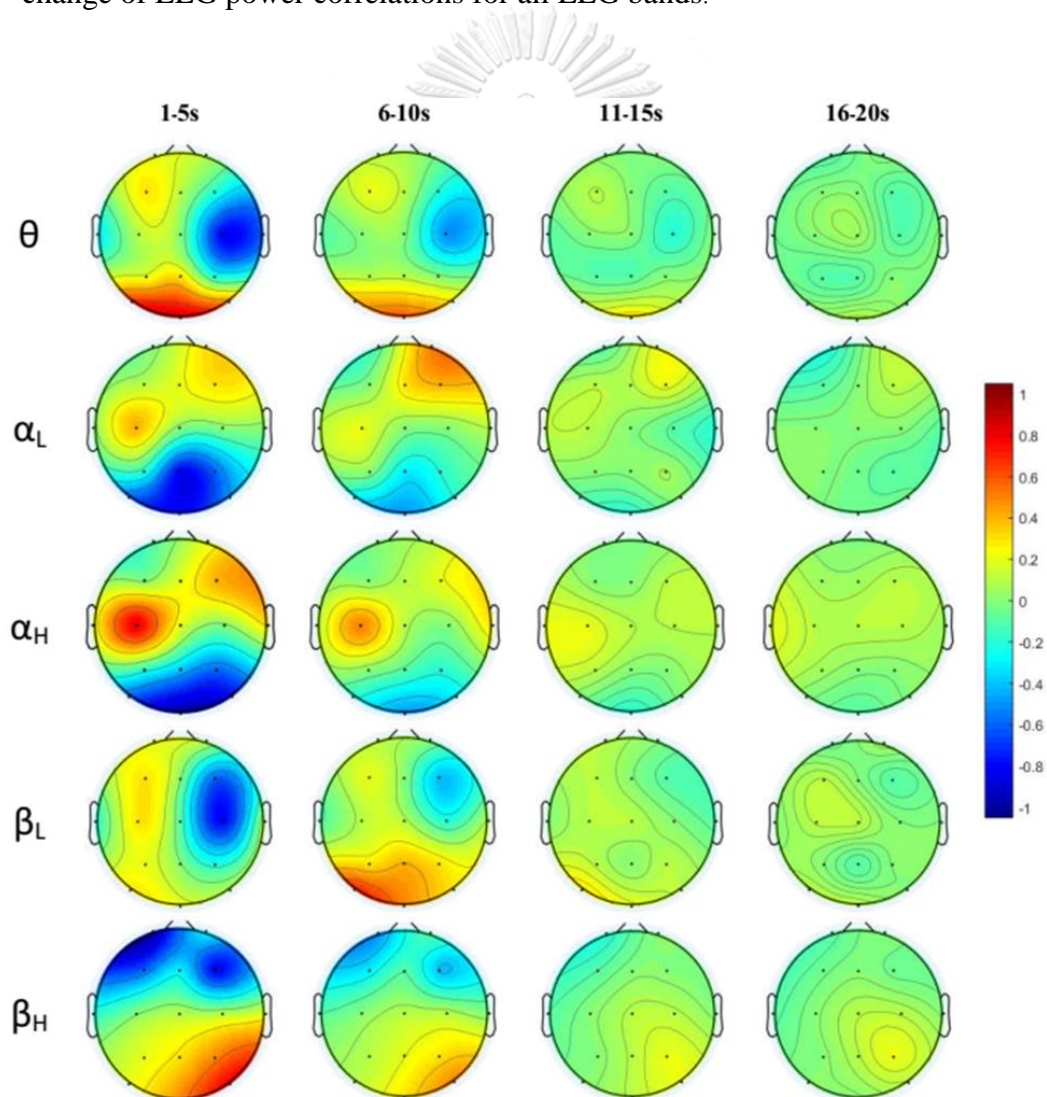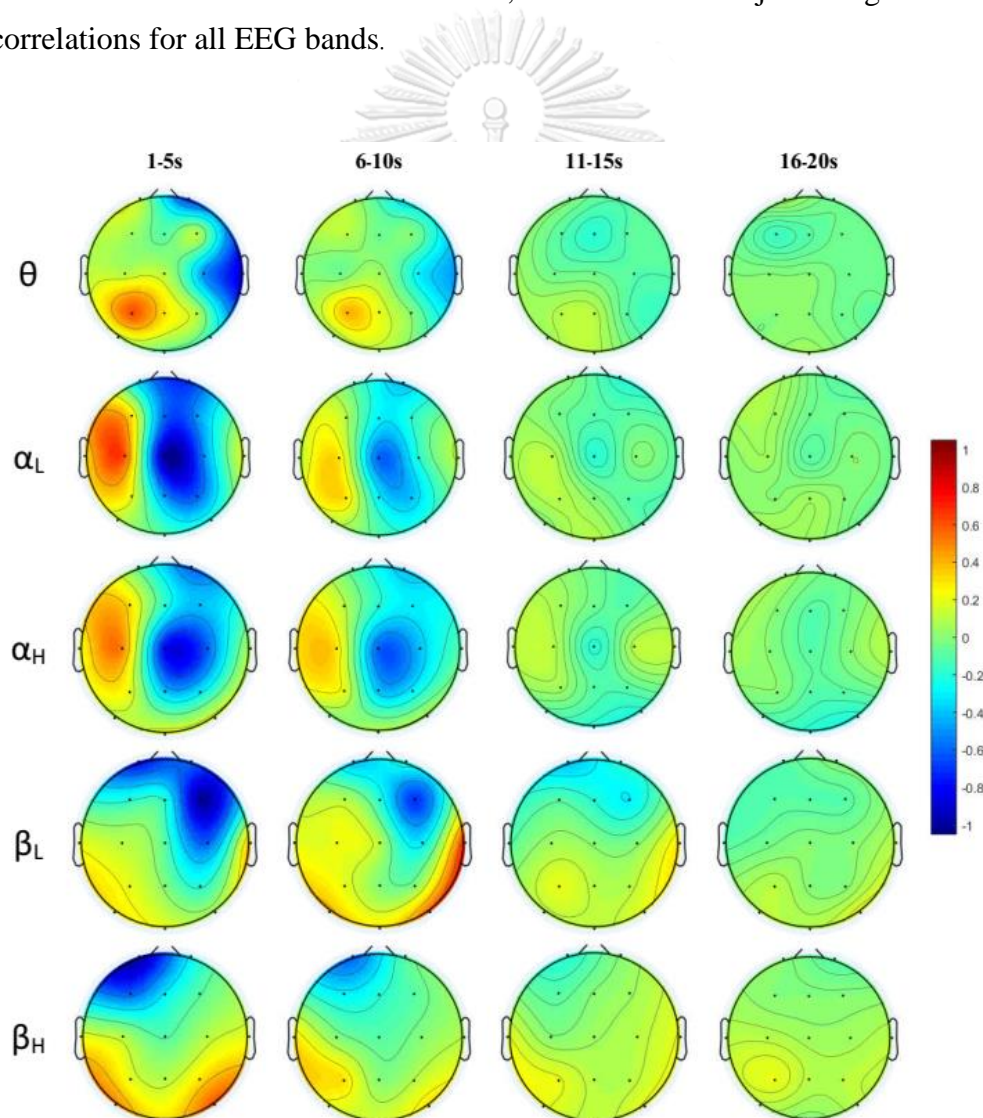*Figure 70   Time-course EEG Power Correlation of Subject 3 (Arousal)*

Chapter 8

Conclusions

EEG-based emotion classification is one of challenging algorithms to make machines or computers respond to their users effectively. The more precision they can predict, the better interaction they can provide. Emotional state regulation primarily involves many brain's functions, especially in limbic and prefrontal cortex area. Electroencephalogram (EEG) signals, derived from brain activities, are associated with many cognitive functions and then propagate though a complex hierarchy of neuron cells. Irrefutably, EEG-based emotion classification requires sophisticated algorithms that can learn complicated EEG feature correlations with the cognitive states of emotions.

This research study fully investigates the feasibility of utilizing DLNs to enhance accuracy performance of the EEG-based emotion classification. The DLN has potential to perform hierarchical feature learning to discover relevant coherences among its inputs. Features at its high-level can be investigated from compositions of features at its low-level with greedy layer-wise unsupervised pre-training. This research study implements EEG-based emotion classifications with two approach: Static DLN and Dynamic DLN-HMM.

.In static DLN approach, the proposed EEG-based emotion recognition is implemented a stack of three autoencoders and two softmax classifiers. This approach employs the concept of PCA is to reduce the number of input features. Moreover, it exploits the concept of CSA to handle the non-stationary effect of EEG signals. Furthermore, this dissertation examines the effects of temporal neural dynamics of emotions and then focuses on relevant state transitions of DLN's high-level features by utilizing Hidden Markov Models (HMMs). The HMM is capable of learning relevant transitions of high-level features of the DLN. From experimental results, the EEG-based Emotion Classification system with hybrid DLN-HMM has better accuracy performance compared with using only DLNs

In dynamic DLN-HMM approach, the proposed EEG-based emotion classification uses the temporal neural dynamics of emotion to enhance the accuracy performance for valence and arousal classifications. The DLNs still provide hierarchical feature learning with greedy layer-wise unsupervised pre-training. The HMMs are responsible for learning all relevant state transitions of top-level DLN features. To classify emotion, the HMMs use the posterior probabilities of state

sequence and then classify the emotion, which has the most posterior probability among emotion classes.

From experimental results, the optimal number of hidden states in HMM for our proposed system is 7 states. However, when the number of hidden states is increased from 5 states to 7 states, the improvement is relatively smaller compared with another setup from 3 states to 5 states. The optimal number of hidden layer in the DLN is 5 layers for both valence and arousal emotion classifications. About f_frame parameter, when the number of input frames is increased from 3-frame to 5-frame, the improvement of accuracy is significantly better. However, another setup from 5-frame to 7-frame provides mixed outcome of improvement.

This dissertation proposes EEG-based emotion classification systems with two approaches (Static vs Dynamic). We need to investigate the accuracy improvement on the Dynamic approach over the Static approach. Statistical analysis is to examine the difference of these two approaches by paired t-test computation. From experimental results, the dynamic DLN+HMM approach provides better accuracy performance than the static DLN approach in both valence and arousal classifications. The optimal system configuration of our proposed method is "n_state HMM = 5", "m_layer DLN = 5" and "f_frame Input Window = 7". With this configuration, the average of accuracy for valence classification improves from 60.07% to 64.18% and those of arousal classification improves from 59.83% to 62.98%.

One of purposes of this dissertation is to obtain understandings about temporal neural dynamics of emotion processes. This research study seeks to determine relevant patterns of EEG power correlations of particular emotional states for individual subjects. The results of EEG power correlation analyses on individual subjects showed that subjects provided their own unique EEG power correlation. Moreover, this research further performs time-course analyses of EEG power correlations in order to discover common patterns of EEG power changes of particular emotional states for individual subjects. Examples of time-course EEG power correlation reveal that there are significant power synchronization and desynchronization of particular EEG bands at specific locations, especially during 10 sec after emotion stimulus. However, there are no significant change of EEG power correlations for all EEG bands after that period.

From the results of time-course analyses, there were significant changes of EEG power correlations during the first 10-sec period after stimulus. One of possible solutions to improve the accuracy of our proposed system is the reduction of number of input features in data sequence from 20 seconds to 10 seconds. However, there are small number of data samples in our dataset for training DLNs. This solution will induce another problem of overfitting.

About our future works, we plan to replace the HMMs with Long Short-Term Memory (LSTM) networks to learn high-level features derived from the DLNs. The LSTM network is a type of Recurrent Neural Network (RNN) architecture. The RNN provides more complex sequential state learning than Hidden Markov Model. The RNN has capability of learning sequential state process from historic information in past but the HMM employs the information of its previous state only. In other words, HMM is constructed with an inference model based on assumptions of Markov process, meaning that the current state solely depends on its previous state only. On the other hands, RNNs implement internal states as their memory to process sequences of inputs and they use their memory to learn sequential process from historical information, not only from current state information. The RNN has a major problem of vanishing gradient for learning long state sequences. However, the LSTM network is capable of learning long-term dependencies because it provides a special cell gate helping to preserve errors during its backpropagation. This mechanism of maintaining constant errors allows LSTM networks to learn long-term dependencies of input feature sequences.

One of LSTM's limitations is lack of efficiency to perform sequential state learning with small number of datasets. LSTM has the capability of learning state sequences from long-term historic information. The complexity computation of such long-term dependency is considerably difficult to learn from small number of data. The algorithm requires sufficient amount of training datasets to learn relevant sequential state transitions. Therefore, in our future works, we need to collect more EEG data during emotion stimulus on the same subject group.

REFERENCES

1. Akram, F., et al. *A novel P300-based BCI system for words typing.* in *2013 International Winter Workshop on Brain-Computer Interface (BCI).* 2013. IEEE.

2. Carlson, T., J.d.R.J.I.R. Millan, and A. Magazine, *Brain-controlled wheelchairs: a robotic architecture.* 2013. **20**(1): p. 65-73.

3. Lee, H., et al. *Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations.* in *Proceedings of the 26th annual international conference on machine learning.* 2009. ACM.

4. Socher, R., et al. *Convolutional-recursive deep learning for 3d object classification.* in *Advances in neural information processing systems.* 2012.

5. Collobert, R. and J. Weston. *A unified architecture for natural language processing: Deep neural networks with multitask learning.* in *Proceedings of the 25th international conference on Machine learning.* 2008. ACM.

6. Hinton, G., et al., *Deep neural networks for acoustic modeling in speech recognition.* 2012. **29**.

7. Hinton, G.E., S. Osindero, and Y.-W.J.N.c. Teh, *A fast learning algorithm for deep belief nets.* 2006. **18**(7): p. 1527-1554.

8. Bengio, Y., et al. *Greedy layer-wise training of deep networks.* in *Advances in neural information processing systems.* 2007.

9. Wulsin, D., et al., *Modeling electroencephalography waveforms with semi-supervised deep belief nets: fast classification and anomaly measurement.* 2011. **8**(3): p. 036015.

10. Längkvist, M., L. Karlsson, and A.J.A.i.A.N.S. Loutfi, *Sleep stage classification using unsupervised feature learning.* 2012. **2012**: p. 5.

11. Bengio, Y.J.F. and t.i.M. Learning, *Learning deep architectures for AI.* 2009. **2**(1): p. 1-127.

12. Lee, T.S. and D.J.J.A. Mumford, *Hierarchical Bayesian inference in the visual cortex.* 2003. **20**(7): p. 1434-1448.

13. Ko, B.J.s., *A brief review of facial emotion recognition based on visual information.* 2018. **18**(2): p. 401.

14. Walecki, R., et al., *Deep structured learning for facial expression intensity estimation.* 2017. **259**: p. 143-154.

15. Kleinsmith, A. and N.J.I.T.o.A.C. Bianchi-Berthouze, *Affective body expression perception and recognition: A survey.* 2012. **4**(1): p. 15-33.

16. Kapur, A., et al. *Gesture-based affective computing on motion capture data.* in *International conference on affective computing and intelligent interaction.* 2005. Springer.

17. Glowinski, D., et al. *Technique for automatic emotion recognition by body gesture analysis.* in *2008 IEEE Computer Society Conference on Computer*

Vision and Pattern Recognition Workshops. 2008. IEEE.

18.   De Silva, P.R., A. Kleinsmith, and N. Bianchi-Berthouze. *Towards unsupervised detection of affective body posture nuances.* in *International Conference on Affective Computing and Intelligent Interaction.* 2005. Springer.

19.   Hosseini, Z., S.M. Ahadi, and N. Faraji. *Speech emotion classification via a modified gaussian mixture model approach.* in *7'th International Symposium on Telecommunications (IST'2014).* 2014. IEEE.

20.   Li, L., et al. *Hybrid Deep Neural Network--Hidden Markov Model (DNN-HMM) Based Speech Emotion Recognition.* in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction.* 2013. IEEE.

21.   Valenza, G., et al., *Revealing real-time emotional responses: a personalized assessment based on heartbeat dynamics.* 2014. **4**: p. 4998.

22.   Cheng, Z., et al. *A novel ECG-based real-time detection method of negative emotions in wearable applications.* in *2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC).* 2017. IEEE.

23.   Wu, C.-K., P.-C. Chung, and C.-J.J.I.T.o.A.C. Wang, *Representative segment-based emotion analysis and classification with automatic respiration signal segmentation.* 2012. **3**(4): p. 482-495.

24.   Mirmohamadsadeghi, L., A. Yazdani, and J.-M. Vesin. *Using cardio-respiratory signals to recognize emotions elicited by watching music video clips.* in *2016 IEEE 18th International Workshop on Multimedia Signal Processing (MMSP).* 2016. IEEE.

25.   Liu, M., et al. *Human emotion recognition based on galvanic skin response signal feature selection and svm.* in *2016 International Conference on Smart City and Systems Engineering (ICSCSE).* 2016. IEEE.

26.   Goshvarpour, A., A. Abbasi, and A.J.b.j. Goshvarpour, *An accurate emotion recognition system using ECG and GSR signals and matching pursuit method.* 2017. **40**(6): p. 355-368.

27.   Gamon, D. and A.D. Bragdon, *Building mental muscle: Conditioning exercises for the six intelligence zones.* 1998: Barnes & Noble New York.

28.   Klem, G.H., et al., *The ten-twenty electrode system of the International Federation.* 1999. **52**(3): p. 3-6.

29.   [cited 2019 April 13]; Available from: https://en.wikipedia.org/wiki/Electroencephalography.

30.   Ekman, P., W.V. Friesen, and P. Ellsworth, *Emotion in the human face: Guidelines for research and an integration of findings.* Vol. 11. 2013: Elsevier.

31.   James, W., *The physical basis of emotion.* 1994.

32.   McDougall, W., *An introduction to social psychology.* 2015: Psychology Press.

33.   Mowrer, O., *Learning theory and behavior.* 1960.

34.   Russell, J.A.J.J.o.p. and s. psychology, *A circumplex model of affect.* 1980. **39**(6): p.

1161.

35.   Horlings, R., D. Datcu, and L.J. Rothkrantz. *Emotion recognition using brain activity.* in *Proceedings of the 9th international conference on computer systems and technologies and workshop for PhD students in computing.* 2008. ACM.

36.   Schmidhuber, J.J.N.n., *Deep learning in neural networks: An overview.* 2015. **61**: p. 85-117.

37.   Jung, T.-P., et al., *Removing electroencephalographic artifacts by blind source separation.* 2000. **37**(2): p. 163-178.

38.   Rabiner, L.R.J.P.o.t.I., *A tutorial on hidden Markov models and selected applications in speech recognition.* 1989. **77**(2): p. 257-286.

39.   Lindquist, K.A., et al., *The brain basis of emotion: a meta-analytic review.* 2012. **35**(3): p. 121-143.

40.   Waugh, C.E., E.Z. Shing, and B.M.J.E.R. Avery, *Temporal dynamics of emotional processing in the brain.* 2015. **7**(4): p. 323-329.

41.   Waugh, C.E., et al., *The role of the medial frontal cortex in the maintenance of emotional states.* 2014. **9**(12): p. 2001-2009.

42.   Goldin, P.R., et al., *The neural bases of emotion regulation: reappraisal and suppression of negative emotion.* 2008. **63**(6): p. 577-586.

43.   Jatupaiboon, N., S. Pan-ngum, and P.J.T.S.W.J. Israsena, *Real-time EEG-based happiness detection system.* 2013. **2013**.

44.   Jie, X., et al., *Emotion recognition based on the sample entropy of EEG.* 2014. **24**(1): p. 1185-1192.

45.   Lin, Y.-P., et al., *EEG-based emotion recognition in music listening.* 2010. **57**(7): p. 1798-1806.

46.   Lin, Y.-P., Y.-H. Yang, and T.-P.J.F.i.n. Jung, *Fusion of electroencephalographic dynamics and musical contents for estimating emotional responses in music listening.* 2014. **8**: p. 94.

47.   Liu, Y. and O. Sourina, *Real-time subject-dependent EEG-based emotion recognition algorithm*, in *Transactions on Computational Science XXIII.* 2014, Springer. p. 199-223.

48.   Nie, D., et al. *EEG-based emotion recognition during watching movies.* in *2011 5th International IEEE/EMBS Conference on Neural Engineering.* 2011. IEEE.

49.   Wang, X.-W., D. Nie, and B.-L. Lu. *EEG-based emotion recognition using frequency domain features and support vector machines.* in *International conference on neural information processing.* 2011. Springer.

50.   Koelstra, S., et al. *Single trial classification of EEG and peripheral physiological signals for recognition of emotions induced by music videos.* in *International Conference on Brain Informatics.* 2010. Springer.

51.   Soleymani, M., et al., *A multimodal database for affect recognition and implicit tagging.* 2011. **3**(1): p. 42-55.

52. Huang, D., et al. *Asymmetric spatial pattern for EEG-based emotion detection.* in *The 2012 International Joint Conference on Neural Networks (IJCNN).* 2012. IEEE.

53. Liu, Y.-H., et al. *Single-trial EEG-based emotion recognition using kernel Eigen-emotion pattern and adaptive support vector machine.* in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* 2013. IEEE.

54. Zheng, W.-L., et al. *EEG-based emotion classification using deep belief networks.* in *2014 IEEE International Conference on Multimedia and Expo (ICME).* 2014. IEEE.

55. Gao, Y., H.J. Lee, and R.M. Mehmood. *Deep learninig of EEG signals for emotion recognition.* in *2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW).* 2015. IEEE.

56. Yang, Y., et al., *EEG-based emotion recognition using hierarchical network with subnetwork nodes.* 2017. **10**(2): p. 408-419.

57. Vincent, P., et al., *Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion.* 2010. **11**(Dec): p. 3371-3408.

58. Koelstra, S., et al., *Deap: A database for emotion analysis; using physiological signals.* 2011. **3**(1): p. 18-31.

59. Morris, J.D.J.J.o.a.r., *Observations: SAM: the Self-Assessment Manikin; an efficient cross-cultural measurement of emotional response.* 1995. **35**(6): p. 63-68.

60. Spüler, M., W. Rosenstiel, and M.J.E.J.o.A.i.S.P. Bogdan, *Principal component based covariate shift adaption to reduce non-stationarity in a MEG-based brain-computer interface.* 2012. **2012**(1): p. 129.

61. Jolliffe, I., *Principal component analysis.* 2011: Springer.

62. Baldi, P. and K.J.N.n. Hornik, *Neural networks and principal component analysis: Learning from examples without local minima.* 1989. **2**(1): p. 53-58.

63. Chang, C.-C., C.-J.J.A.t.o.i.s. Lin, and technology, *LIBSVM: A library for support vector machines.* 2011. **2**(3): p. 27.

64. Li, K., et al. *Affective state recognition from EEG with deep belief networks.* in *2013 IEEE International Conference on Bioinformatics and Biomedicine.* 2013. IEEE.

65. Chung, S.Y. and H.J. Yoon. *Affective classification using Bayesian classifier and supervised learning.* in *2012 12th International Conference on Control, Automation and Systems.* 2012. IEEE.

66. Lotte, F. and C. Guan. *Learning from other subjects helps reducing brain-computer interface calibration time.* in *2010 IEEE International conference on acoustics, speech and signal processing.* 2010. IEEE.

67. Samek, W., F.C. Meinecke, and K.-R.J.I.T.o.B.E. Müller, *Transferring subspaces*

between subjects in brain-computer interfacing. 2013. **60**(8): p. 2289-2298.

68. October 17th, 2018; Available from: https://hmmlearn.readthedocs.io/en/latest/changelog.html#version-0-2-1.

69. [cited June 12, 2019; Available from: https://keras.io/getting-started/functional-api-guide/.

70. Boger, Z. and H. Guterman. *Knowledge extraction from artificial neural network models.* in *1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation.* 1997. IEEE.

71. Linoff, G.S. and M.J. Berry, *Data mining techniques: for marketing, sales, and customer relationship management.* 2011: John Wiley & Sons.

72. Blum, A., *Neural networks in C++: an object-oriented framework for building connectionist systems.* 1992: John Wiley & Sons, Inc.

73. Hochreiter, S., et al., *Gradient flow in recurrent nets: the difficulty of learning long-term dependencies.* 2001, A field guide to dynamical recurrent neural networks. IEEE Press.

74. Hochreiter, S.J.I.J.o.U., Fuzziness and K.-B. Systems, *The vanishing gradient problem during learning recurrent neural nets and problem solutions.* 1998. **6**(02): p. 107-116.

75. Pasini, A.J.J.o.t.d., *Artificial neural networks for small dataset analysis.* 2015. **7**(5): p. 953.

76. Aftanas, L.I., et al., *Analysis of evoked EEG synchronization and desynchronization in conditions of emotional activation in humans: temporal and topographic characteristics.* 2004. **34**(8): p. 859-867.

77. Barry, R.J., et al., *EEG differences between eyes-closed and eyes-open resting conditions.* 2007. **118**(12): p. 2765-2773.

78. Du, R. and H.J. Lee. *Power spectral performance analysis of EEG during emotional auditory experiment.* in *2014 International Conference on Audio, Language and Image Processing.* 2014. IEEE.

79. Eijlers, E., A. Smidts, and M.A.J.P.o. Boksem, *Implicit measurement of emotional experience and its dynamics.* 2019. **14**(2): p. e0211496.

80. Hu, X., et al., *EEG correlates of ten positive emotions.* 2017. **11**: p. 26.

81. Jäncke, L., et al., *Time course of EEG oscillations during repeated listening of a well-known aria.* 2015. **9**: p. 401.

82. Kroupi, E., et al. *EEG correlates during video quality perception.* in *2014 22nd European signal processing conference (EUSIPCO).* 2014. IEEE.

83. Lin, Y.-P., et al., *Electroencephalographic dynamics of musical emotion perception revealed by independent spectral components.* 2010. **21**(6): p. 410-415.

84. Shemyakina, N. and S.J.H.P. Dan'ko, *Changes in the power and coherence of the β 2 EEG band in subjects performing creative tasks using emotionally significant and emotionally neutral words.* 2007. **33**(1): p. 20-26.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

APPENDIX

Table 21   Valence Accuracy of Data Size Comparison (n_state)

Crossvalidation (10-fold)	n=3		n=5		n=7	
	N=1200	**N=2400**	**N=1200**	**N=2400**	**N=1200**	**N=2400**
#1	58.33%	59.17%	63.33%	63.75%	65.83%	64.17%
#2	63.33%	62.92%	74.17%	74.58%	61.67%	60.83%
#3	54.17%	57.08%	64.17%	63.33%	63.33%	64.58%
#4	62.50%	64.58%	70.83%	69.17%	58.33%	60.83%
#5	66.67%	65.42%	69.17%	70.42%	70.83%	70.42%
#6	64.17%	63.33%	60.83%	60.42%	64.17%	66.67%
#7	69.17%	68.75%	65.83%	66.25%	56.67%	56.67%
#8	66.67%	65.83%	69.17%	70.42%	70.83%	68.75%
#9	65.00%	64.58%	59.17%	61.67%	75.83%	77.50%
#10	66.67%	66.67%	71.67%	70.83%	65.83%	67.50%
Mean	**63.67%**	**63.83%**	**66.83%**	**67.08%**	**65.33%**	**65.79%**
S.D.	**4.47**	**3.47**	**4.94**	**4.67**	**5.91**	**5.85**
p	*0.355381*		*0.268136*		*0.204874*	
p_n	*0.130410*					

p_n : p value of all pairs in n_state variants.

Table 22   Arousal Accuracy of Data Size Comparison (n_state)

Crossvalidation (10-fold)	n=3		n=5		n=7	
	N=1200	**N=2400**	**N=1200**	**N=2400**	**N=1200**	**N=2400**
#1	55.83%	57.08%	56.67%	56.25%	59.17%	60.83%
#2	60.00%	59.58%	61.67%	60.83%	63.33%	64.58%
#3	58.33%	61.25%	59.17%	59.58%	59.17%	60.42%
#4	59.17%	60.42%	60.83%	61.67%	58.33%	56.67%
#5	62.50%	61.67%	65.83%	63.33%	71.67%	70.42%
#6	66.67%	68.33%	67.50%	66.25%	72.50%	71.67%
#7	60.83%	58.75%	61.67%	62.08%	65.83%	63.33%
#8	68.33%	70.83%	69.17%	71.67%	70.83%	72.08%
#9	59.17%	58.33%	61.67%	62.50%	68.33%	70.83%
#10	61.67%	63.33%	67.50%	66.25%	66.67%	68.75%
Mean	**61.25%**	**61.96%**	**63.17%**	**63.04%**	**65.58%**	**65.96%**
S.D.	**3.79**	**4.44**	**4.09**	**4.23**	**5.39**	**5.51**
p	*0.103413*		*0.391648*		*0.258660*	
p_n	*0.141673*					

p_n : p value of all pairs in n_state variants.

Table 23   *Valence Accuracy of Data Size Comparison (m_layer)*

Crossvalidation (10-fold)	m=3		m=5		m=7	
	N=1200	**N=2400**	**N=1200**	**N=2400**	**N=1200**	**N=2400**
#1	59.17%	59.58%	63.33%	63.75%	61.67%	62.92%
#2	68.33%	66.67%	74.17%	74.58%	70.83%	72.08%
#3	61.67%	60.83%	61.27%	63.33%	60.83%	59.58%
#4	68.33%	67.08%	70.83%	69.17%	67.50%	68.33%
#5	63.33%	65.83%	69.17%	70.42%	66.67%	66.67%
#6	57.50%	57.92%	60.83%	60.42%	58.33%	60.42%
#7	59.17%	59.58%	65.83%	66.25%	60.83%	58.75%
#8	68.33%	67.92%	69.17%	70.42%	65.83%	66.25%
#9	60.83%	62.92%	59.17%	61.67%	60.00%	60.83%
#10	65.83%	67.08%	71.67%	70.83%	69.17%	68.75%
Mean	**63.25%**	**63.54%**	**66.83%**	**67.08%**	**64.17%**	**64.46%**
S.D.	**4.20**	**3.80**	**4.94**	**4.67**	**4.34**	**4.57**
p	*0.259464*		*0.268136*		*0.240939*	
p_m	*0.115480*					

p_m : p value of all pairs in m_layer variants.

Table 24   *Arousal Accuracy of Data Size Comparison (m_layer)*

Crossvalidation (10-fold)	m=3		m=5		m=7	
	N=1200	**N=2400**	**N=1200**	**N=2400**	**N=1200**	**N=2400**
#1	54.14%	53.75%	56.67%	56.25%	55.83%	55.42%
#2	56.67%	54.14%	61.67%	60.83%	57.50%	58.33%
#3	60.83%	62.92%	59.17%	59.58%	60.83%	61.67%
#4	55.83%	58.33%	60.83%	61.67%	56.67%	55.83%
#5	63.33%	63.75%	65.83%	63.33%	65.83%	64.17%
#6	64.17%	64.58%	67.50%	66.25%	65.00%	65.83%
#7	60.83%	59.58%	61.67%	62.08%	59.17%	60.42%
#8	67.50%	68.75%	69.17%	71.67%	68.33%	67.08%
#9	60.00%	60.83%	61.67%	62.50%	61.67%	60.42%
#10	63.33%	65.42%	67.50%	66.25%	66.67%	68.75%
Mean	**60.66%**	**61.21%**	**63.17%**	**63.04%**	**61.75%**	**61.79%**
S.D.	**4.16**	**4.85**	**4.09**	**4.23**	**4.49**	**4.59**
p	*0.154582*		*0.391648*		*0.459767*	
p_m	*0.279828*					

p_m : p value of all pairs in m_layer variants.

Table 25   Valence Accuracy of Data Size Comparison (f_frame)

Crossvalidation (10-fold)	f=3		f=5		f=7	
	N=1200	N=2400	N=1200	N=2400	N=1200	N=2400
#1	61.67%	62.92%	60.83%	59.58%	63.33%	63.75%
#2	65.83%	65.83%	71.67%	71.25%	74.17%	74.58%
#3	61.67%	63.33%	66.67%	67.50%	64.17%	63.33%
#4	67.50%	65.83%	72.50%	71.67%	70.83%	69.17%
#5	65.83%	64.58%	71.67%	69.17%	69.17%	70.42%
#6	58.33%	56.67%	55.83%	57.50%	60.83%	60.42%
#7	60.83%	60.42%	61.67%	60.83%	65.83%	66.25%
#8	64.17%	64.58%	67.50%	67.92%	69.17%	70.42%
#9	53.33%	54.17%	56.67%	58.33%	59.17%	61.67%
#10	64.17%	66.67%	68.33%	67.08%	71.67%	70.83%
Mean	**62.33%**	**62.50%**	**65.33%**	**65.08%**	**66.83%**	**67.08%**
S.D.	**4.19**	**4.17**	**6.20**	**5.45**	**4.94**	**4.67**
p	*0.360344*		*0.287349*		*0.268136*	
pf	*0.409727*					

p_f: p value of all pairs in f_frame variants.

Table 26   Arousal Accuracy of Data Size Comparison (f_frame)

Crossvalidation (10-fold)	f=3		f=5		f=7	
	N=1200	N=2400	N=1200	N=2400	N=1200	N=2400
#1	50.83%	52.08%	54.17%	56.25%	56.67%	56.25%
#2	55.83%	54.58%	62.50%	61.67%	61.67%	60.83%
#3	53.33%	54.17%	57.50%	56.67%	59.17%	59.58%
#4	54.17%	53.75%	58.33%	59.58%	60.83%	61.67%
#5	62.50%	64.17%	66.67%	65.83%	65.83%	63.33%
#6	63.33%	63.33%	66.67%	67.08%	67.50%	66.25%
#7	56.67%	57.08%	59.17%	60.42%	61.67%	62.08%
#8	65.83%	67.50%	70.83%	68.75%	69.17%	71.67%
#9	60.00%	62.08%	63.33%	64.58%	61.67%	62.50%
#10	61.67%	61.25%	64.17%	66.67%	67.50%	66.25%
Mean	**58.42%**	**59.00%**	**62.33%**	**62.75%**	**63.17%**	**63.04%**
S.D.	**4.94**	**5.32**	**5.06**	**4.46**	**4.10**	**4.23**
p	*0.063772*		*0.200866*		*0.391648*	
p$_f$	*0.212615*					

p_f : p value of all pairs in f_frame variants.

VITA

NAME   Mr. Suwicha Jirayucharoensak

DATE OF BIRTH   19 September 1974

PLACE OF BIRTH   Bangkok

INSTITUTIONS ATTENDED
-Master of Science in Computer Engineering University of Southern California, Los Angeles, USA. Research Topic "Design and Implementation Asynchronous Pipeline"
-Bachelor of Science in Electrical Engineering Chulalongkorn University, Bangkok, Thailand. Senior Project "Design and Implementation of Voice Mail System"

HOME ADDRESS   37 Soi Chan 32 Yaek 2, Thung Wat Don, Sathorn, Bangkok, Thailand 10120

PUBLICATION
-Jirayucharoensak, S., Israsena, P., Pan-ngum, S., Hemrungrojn, S., & Maes, M. (2019). A game-based neurofeedback training system to enhance cognitive performance in healthy elderly subjects and in patients with amnestic mild cognitive impairment. Clinical interventions in aging, 14, 347.
- Sudnawa, K. K., Chirdkiatgumchai, V., Ruangdaraganon, N., Khongkhatithum, C., Udomsubpayakul, U., Jirayucharoensak, S., & Israsena, P. (2018). Effectiveness of neurofeedback versus medication for attention☐ deficit/hyperactivity disorder. Pediatrics International, 60(9), 828-834.
-Jirayucharoensak, S., Pan-Ngum, S., & Israsena, P. (2014). EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation. The Scientific World Journal, 2014.
-Jirayucharoensak, S., Israsena, P., Pan-ngum, S., & Hemrungrojn, S. (2013, October). Online EEG artifact suppression for neurofeedback training systems. In The 6th 2013 Biomedical Engineering International Conference (pp. 1-5). IEEE.