

**Mood perception model for social robot based on  
Human facial and bodily expression**

JIRAPHAN INTHIAM



## Table of contents

	Page
<b>Table of contents</b>	<b>i</b>
Chapter 1 Introduction	1
1.1 Background	1
1.2 Problem statement	2
1.3 Research propose	3
1.4 Overview of the thesis	4
Chapter 2 Fundamental theory	5
2.1 Kinematic of manipulator	5
2.1.1 Introduction	5
2.1.2 Forward kinematics	6
2.2 Neural Networks for robotics system	8
2.2.1 Introduction	8
2.2.2 Neural Networks	8
2.2.3 Neural Networks and back-propagation	10
2.2.4 Convolutional Neural Networks	14
2.3 Genetic Algorithm	16
2.3.1 Introduction	16
2.3.2 Basic genetic algorithm	16
2.3.3 Crossover	17
2.3.4 Mutation	19
2.3.5 Genotype-phenotype mapping	21
2.3.6 Fitness	21
2.3.7 Selection	22
2.3.8 Termination	23

2.4 Laban movement analysis	23
2.4.1 Introduction	23
2.4.2 LMA-effort	24
2.4.3 LMA-shape	25
2.5 Markov model	27
2.5.1 Markov Chains	27
2.5.2 Hidden Markov Model	31
2.5.3 Viterbi Algorithm	32
2.6 Summary	37
Chapter 3 Methodology	38
3.1 Robot system configuration	38
3.1.1 Hardware configuration	39
3.1.2 RGB-Depth camera (KinectV2)	47
3.2 Study of emotional expression	48
3.2.1 Emotions, Feelings, Moods, and its expression	49
3.2.2 Data collecting process	49
3.2.3 Body states extraction	52
3.3 Mood perception model using HMM	58
3.4 Solving kinematic arm using NNs and backpropagation	63
3.4.1 Kinematic of Seamless robot's arm	63
3.4.2 Neural Networks for solving kinematic problem	64
3.4.2 Robot gesture generation based human emotional expression	65
3.5 Summary	69
Chapter 4 Experimental results	70
4.1 Experiment I: Seamless robot emotional expression	70
4.2 Experiment II: Mood perception system	74

4.2.1 Comparative accuracy of Mood estimation	77
4.2.2 Human-robot emotional interactive task	81
4.3 Summary	83
Chapter 5 Conclusions	84
5.1 Conclusions of the research	84
5.2 Recommendations for future research	87
References	88
Appendix	92
Appendix A: List of predefined command pairing with robot tasks	92
Appendix B: Critical values of the $t$ distribution	93
Appendix C: Publications and Presentations from the Present Research Work	94
List of Figures	95
List of Tables	98
Acknowledgement	100

# Chapter 1

## Introduction

### 1.1 Background

Robot technology is playing an ever greater role in contemporary society. Robots have been used in industrial applications for some time, but now they have a presence in the home as well. To achieve the goal, to develop a Human-Robot Interaction (HRI) system in ways that are natural and easily understood, even by non-roboticists. Companion robots are being designed with several kind of human-like interactive skills such as voice communication, hands-gestures recognition, etc. However, the critical feature of human interactive skills, emotional interaction, is still a huge challenge.

During communication, people exchanges information not only the content from the spoken words, but also emotional through nonverbal messages. Numbers of psychological research also reported a study of emotional expression through communication, facial and bodily expression. The 7-38-55% rule of personal communication was presented [1] , the model study on the relative importance of verbal and nonverbal communications. This rule describes the ratio of elements of personal interaction that consists of 7% (Content), 38% (Tone of voice) and 55% (non-verbal). The work reported that non-verbal elements are particularly essential for communicating feelings and attitude, especially when they are incongruent: if words and body language disagree, and communicator is not talking about their mood or attitude; one tends to believe the body language. The most extensively studied type of non-verbal is one focusing on recognition of an emotional state from facial expression [2], [3]. Additionally, bodily expression behaviour offers a rich source of information about the human emotional states as reported [4], [5]. Moreover, the bodily movement has been suggested as an additional parameter to use together with a facial expression to discriminate between positive and negative moods [6] as well. These

are relevant for the use of physical expressions, facial and body expression to assess human emotional states. Thus endowing the robot with emotional skill will make the flexibility to interact in a more human-oriented way.

Models of emotion have attracted considerable attention in connection with the development of social robot skills. One such model, designed to enhance a robot decision-making system, has been presented in [7]. Another approach, adopted in [8], argues for designing a robot's behaviour selection system based on the expression of emotion in the robot itself. Common to both of these approaches to modeling emotion is an emphasis on situation and environment. However, the interaction between humans and robots should be taken into account in any model of emotion related to a robot's behaviour. In a study of junior high school students, the effect of collaborative learning involving such interaction was shown to promote more learning than collaboration between humans without the presence of a robot [9]. The models discussed in the literature on robots and emotion have been examined here to provide a foundation for the model presented in this thesis.

With this observation in mind, Hayashi's laboratory has been focused on studying and developing the emotional interactive skill, in order to enhance the new ability of a social robot. The ultimate goal of this work is not only to enhance a new robotic perception, but also the ability of a robot to communicate information that humans can perceive. With this 'full duplex' capability, HRI will become truly effective.

## **1.2 Problem statement**

As noted above, emotional skill is an essential key to improve interaction skills for social-robot. The extremely research topic, have been studied in emotional assessment is an emotional perception based on facial expression. However, there is an abandoned feature, bodily behaviour, offers a rich source of information about the human emotional states. Thus, bodily movements should be studied for a better understanding in the way that human express their feelings. These

knowledge could be used as basement for developing emotional recognition for robotics system. On top of that, there are possibility of using bodily information to generate robot behaviour capable of revealing its emotional state. The robot with physical expression capabilities could communicate more successfully and provide a better suited for interaction with humans than a bodiless computer program. Consequently, all problems will be considered and developed by the proposed system in this thesis.

### **1.3 Research propose**

For several decades, body movements have been largely disregarded in emotion research although they make an essential contribution to emotion recognition. Our aim in the system described here is to the development of improved emotional interactive skills in robotic systems based on study of human-physical expression. The general objective of this project is to study a behaviour of physical-emotional expression of human in order to create a new emotional interactive skill for the robotics system. The specific objectives are established as follows:

1. To study emotional expression through human physical behaviour and reducing a complexity of bodily movement into simple state of expression based on Laban movement analysis.
2. To create an emotional expression generation model, that can express emotional state based on studied data of human expression.
3. To create a new model to estimate a human mood with a new essential idea by using bodily expression as additional channel for assessing internal feeling of the interlocutor. The model is desired to estimate the simplest mood states, positive or negative state, since these two simple state provide enough information for the robot to choose the appropriate reaction when interacting with humans.
4. To verify the used of bodily expression in emotional interactive system for the social robot, and confirm the effectiveness of the proposed system with experimental results.



## 1.4 Overview of the thesis

The thesis consists of five chapters covering the background history, research objectives, modeling of the emotional expression generation system, and mood perception model.

Chapter 2 describes the fundamental theories for Seamless robot system, including a Laban Movement Analysis (LMA), Kinematics modeling for controlling the robot movement, Hidden Markov Model and Face emotional recognition model based Convolution Neural Network (CNN), which are used to construct and develop the emotional interactive system.

Chapter 3 shows the system structure of the Seamless robot; it is divided as the system configuration and overview of the software. At the end of this chapter, the computation of robotic arm posture is explained to perform the sequential movement.

Chapter 4 the proposed system is thoroughly described in this chapter consist with the following process: First, the data collecting process is desired for collect realistic expressional data from the participants when they were expressing their feeling in front of the other. Then, the expression data were analysed regard to theory of human movement. The result of the studied data has been used as a basement for creating an emotional interactive system. Finally, Seamless robot has been implemented to show the effectiveness of the proposed system.

In the end, chapter 5 is a conclusion that summarizes the thesis, and suggestion the future works, which has been mentioned in order to develop and improve the emotional interactive system in the future.

## Chapter 2

### Fundamental theory

#### 2.1 Kinematic of manipulator

##### 2.1.1 Introduction

Definition: “A robot is a reprogrammable, multifunctional manipulator designed to move material parts, tools, or specialized devices through variable programmed motions for the performance of a variety of tasks. [10]” That is the programmability, which gives a robot its utility and adaptability. The first successful applications of robot manipulators generally involved transfer of parts in the factories, or that could be programmed to execute the sequential movements, such as moving to position A, closing a gripper, moving to position B, and opening a gripper, etc., and more complex applications, such as welding, grinding, deburring, and assembly that require not only more complex motion but also some form of the external sensing such as vision, tactile, distance, or force sensing, due to increase accurately the interaction of the robot with its environment

Kinematics is the science of motion that studies the body movements without considering the forces or moments, size, shape and weight. The formulation of the suitable kinematics model for a robot mechanism is very important for analyzing the part movements of industrial manipulators. There are commonly two different spaces used in kinematics modeling of manipulators such as Cartesian space and Quaternion space. There are many ways to represent rotation, including the following: Euler angle, Roll-Pitch-Yaw angles, Cayley Klein parameters, Pauli spin matrices, orthonormal matrix and Hamilton's quaternions. The general transformation between two joints that requires four parameters is Denavit & Hartenberg [11]. These parameters known as the D-H parameters have become the standard for describing robot kinematics models. For representing the rotation and translation in a compact form of transformation vector that can

be determined as a Dual quaternion. For example, if the body is represented nine elements in the homogeneous transformations, the dual quaternion can reduce the number of elements to four parameters. It is accepted as an advantage in terms of computational robustness and storage efficiency for dealing with the kinematics of robot chains.

The robot kinematics can be divided into forward kinematics and inverse kinematics. Forward kinematics problem is static geometrical problem in solving the equations. Because the typical manipulator will be able to sense its own position by using internal sensors that can be directly measured the joint angles, and then it is determined the position and orientation of the end effector or tool frame, which relative to the based frame. But the inverse end effector kinematics problem is much more difficult problem than forward kinematics. Singularities and nonlinearities that make the problem more difficult to solve the solution of the inverse kinematics problem. Relationship between forward and inverse kinematics is shown in Figure 2-1.

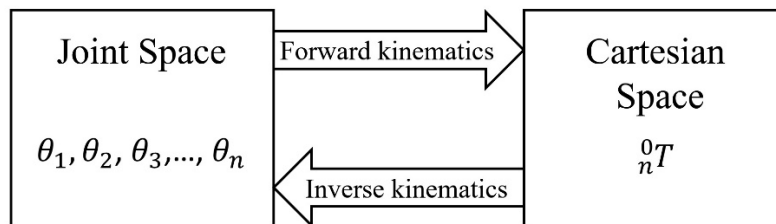


Figure 2-1 Schematic representation of forward and inverse kinematics.

### 2.1.2 Forward kinematics

A robotic manipulator is composed of serial links which are connected to each other revolute or prismatic joints from the base frame to the end effector or tool frame. Determining the position and orientation of the end-effector in terms of the joint variables called as forward kinematics. One method will be used for a suitable kinematics model that is Denavit-Hartenberg method. This method uses four parameters (D-H parameters) is the most common method for describing the robot kinematics. The parameters are the link length  $a_{i-1}$ , link twist  $\alpha_{i-1}$ , link offset  $d_i$ , and joint angle  $\theta_i$ . In Figure 2-2 that shows coordinate frame assignment for general

manipulator. As description in the figure, the distance from  $z_{i-1}$  to  $z_i$ , along  $X_i$  is assigned as  $a_{i-1}$ . the angle between  $Z_{i-1}$  to  $Z_i$  about  $X_i$  is assigned as  $\alpha_{i-1}$ , the distance from  $X_{i-1}$  to  $X_i$  along  $Z_i$  is assigned as  $d_i$  the angle between  $X_{i-1}$  to  $X_i$  about  $Z_i$  is assigned as  $\theta_i$ . Thus, the general transformation matrix  ${}^{i-1}_i T$  for a single link can be obtained as follow by Equation (2-1)

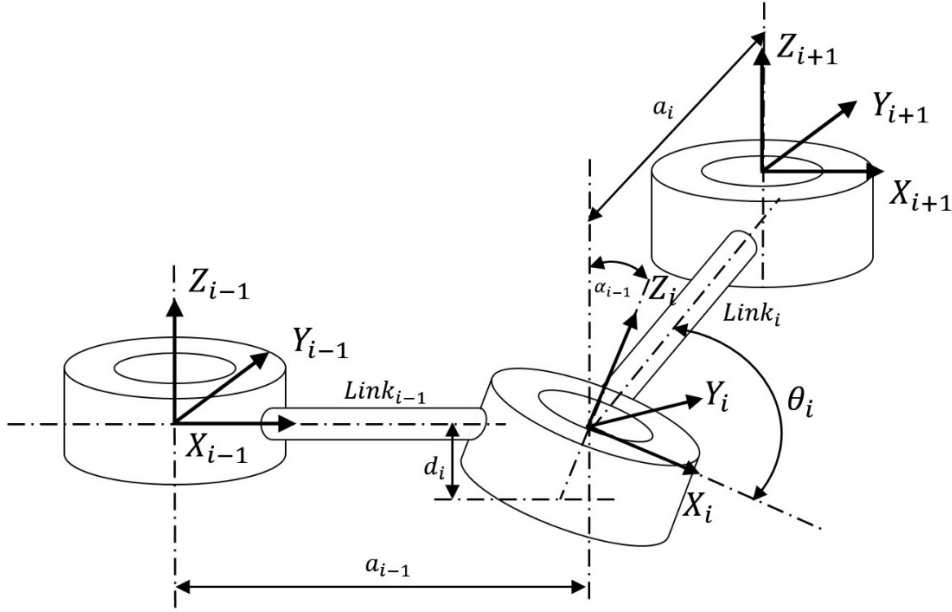


Figure 2-2 Denavit-Hartenberg frame assignment

$${}^{i-1}_i T = R_x(\alpha_{i-1})D_x(a_{i-1})R_z(\theta_i)D_z(d_i)$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & c\alpha_{i-1} & -s\alpha_{i-1} & 0 \\ 0 & s\alpha_{i-1} & c\alpha_{i-1} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & \alpha_{i-1} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c\theta_i & -s\theta_i & 0 & 0 \\ s\theta_i & c\theta_i & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} c\theta_i & -s\theta_i & 0 & a_{i-1} \\ s\theta_i c\alpha_{i-1} & c\theta_i c\alpha_{i-1} & -s\alpha_{i-1} & -s\alpha_{i-1} d_i \\ s\theta_i s\alpha_{i-1} & c\theta_i s\alpha_{i-1} & c\alpha_{i-1} & c\alpha_{i-1} d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2-1)$$

The forward kinematics of the end-effector with respect to the based frame determines by multiply all transition matrices as expressed by Equation (2-2).

$${}_{end-effector}^{base}T = {}_1^0T {}_2^1T {}_3^2T \dots {}_n^{n-1}T$$

$$= \begin{bmatrix} r_{11} & r_{12} & r_{13} & p_x \\ r_{21} & r_{22} & r_{23} & p_y \\ r_{31} & r_{32} & r_{33} & p_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2-2)$$

$r_{ij}$  define as the rotational element of the transformation matrix ( $i$  and  $j = 1, 2$  and  $3$ ).

The elements of the position vector denote with  $p_x$ ,  $p_y$  and  $p_z$ .

## 2.2 Neural Networks for robotics system

### 2.2.1 Introduction

Nowadays, Neural Networks (NNs), or more precisely Artificial Neural Networks (ANNs) is the most powerful tools have been used in difference fields of research since NNs has the advantage capability to model complex relationships between inputs and outputs or to find patterns in data. NNs is applied in many application areas such as control (vehicle control, robot manipulator control), decision making, pattern recognition (face identification, object recognition), sequence recognition (gesture, speech recognition). With this utilities, NNs has become an interesting topic for the researchers in robotics field.

### 2.2.2 Neural Networks

NN is a computing system inspired by biological neural network in animal brain that solving the task by using mathematic to computing from the given example, generally without being programmed with any task-specific rules. The NNs is constructed from 3 type of layer as showed in Figure 2-3 consist of 1) Input layer-initial data for the neural network, 2) Hidden layer-

intermediate layer between input and output layer and place where all the computation is done, and 3) Output layer - produce the result for given inputs.

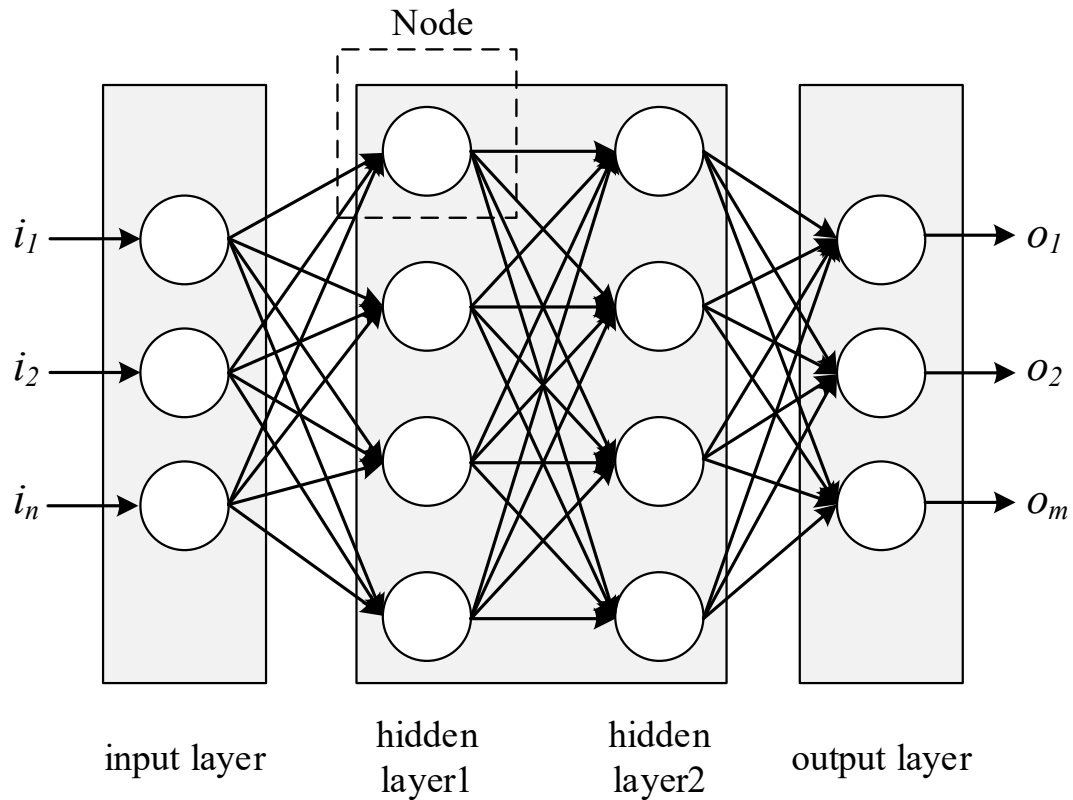


Figure 2-3 ANN architecture

The layers are made of nodes, which each node is a place where computation happens. The inputs from the data with a set of weights, and biases, that either amplify or dampen that input, thereby assigning significance to inputs with regard to the task the algorithm is trying to learn. These input-weight products are summed and then the sum is passed through a node's so-called activation function, to determine whether and to what extent that signal should progress further through the network to affect the ultimate outcome. The diagram of one node is shown in Figure 2-4

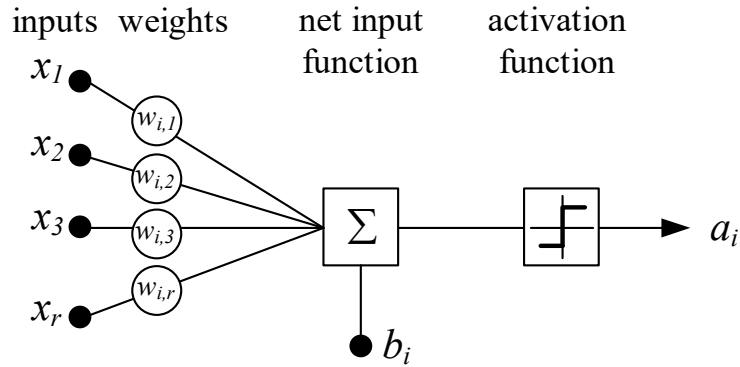


Figure 2-4 Single node diagram

The components in node as represented in Figure2-4 are:

1.  $x_j$ : Inputs to node. These can either be data from input layer or an intermediate output layer from one of hidden layers.

2.  $b_i$ : Biases unit. This is a constant value added to the input of the activation function. Bias helps in controlling the value at which activation function will trigger.

3.  $w_{i,j}$ : Weights refers to the strength of a connection between two nodes which is modified by using a learning rule in order to minimize total loss function to fit the desired model

4.  $z_i$ : Outputs of the single node which is calculated as:

$$z_i = g\left(b_i + \sum_{j=0}^r w_{i,j}x_j\right) \quad (2-3)$$

where  $g$  is known an activation function which can be a Gaussian function, logistic function, hyperbolic function or even a linear function in a simple cases. This makes Neural Network extremely flexible and imparts the capability to estimate complex non-linear relationships in data.

### 2.2.3 Neural Networks and back-propagation

This section is to explain how neural networks work to fit the desired model. The learning mechanism of NN will be explained with the simple abstraction, focusing only on the concepts.

A simplest abstraction can be presented as a black box with 2 method learn and predict as shown in Figure2-5

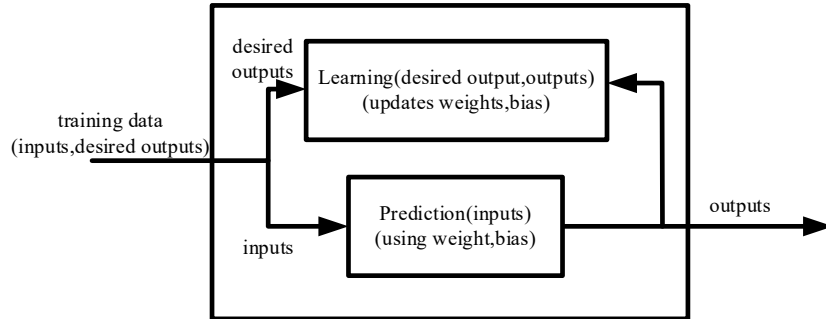


Figure 2-5 Neural network as a black box

The process takes the training data consist with inputs, and desired output. The input data is fed to the prediction process, using weights and biases in order to calculate the output. The difference between calculated output and desired output will be used for updating the weights and biases, to make the calculated output get as close as possible from the desired output. The overall process of NN and back-propagation is described by the basic structure of 2-layer NN in Figure 2-6:

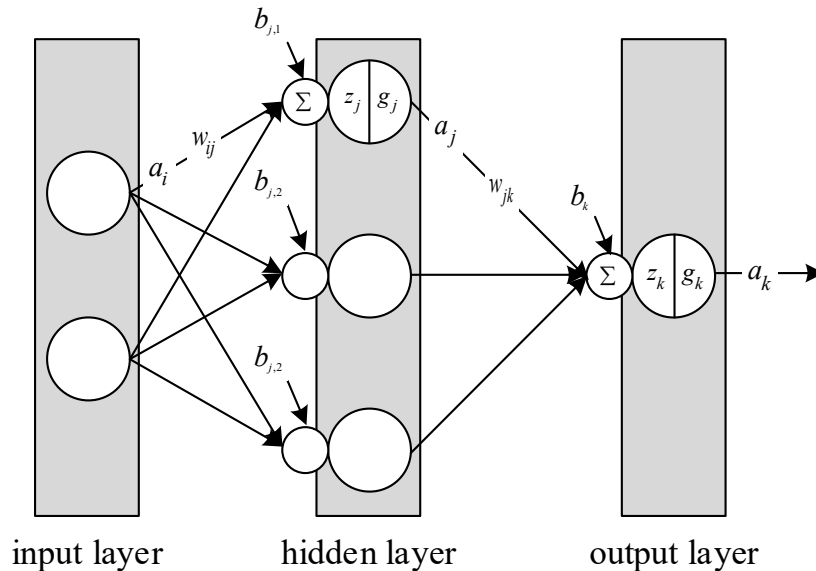


Figure 2-6 Diagram of multi-layer NN



From Figure2-5, associate index  $i$  with the input later, index  $j$  with hidden layer and index  $k$  with the output layer. The process of NN is described by following step:

**Step 1: Model Initialization**, the set of weights ( $w$ ) and bias ( $b$ ) in every internal layer are random initialized.

**Step 2: Forward propagation**, the second step is to check its performance. The training data (input) is passed through the network layer and calculate the total net input to hidden node, or called pre-activations value ( $z$ ) is computed by Equation (2-4), then the output of each node, activated value ( $a$ ) is given by activating  $z$  with Equation (2-5).

$$z_j = b_j + \sum_i a_i w_{ij} \quad (2-4)$$

$$a_j = g_j(z_j) \quad (2-5)$$

The activated output from the previous layer is used as the input data for next layer, so the output at the final layer of diagram  $a_k$  in Figure(2-5) by the given input  $a_i$  is thus obtained by Equation (2-6).

$$a_k = g_k(b_k + \sum_j a_j w_{jk}) \quad (2-6)$$

Note: the value of both pre-activation and activated for all layer are keep tracked in order to calculating in back-propagation process.

**Step 3: Loss function**, Loss function is a method of evaluating how well algorithm models the dataset. If the prediction are totally off, loss function will output high number. If they're pretty good, its output a lower number. At this step, we have the output from the randomly NN model. However, this model need to learn and update itself regard to the desired output. So we can design our loss function with the simplest idea by the difference between our output and desired output, or we can define the loss function to be the sum of the Squared Error, expressed by Equation (2-7).

$$loss = \frac{1}{2} \sum_{k \in K} (a_k - t_k)^2 \quad (2-7)$$

**Step 4: Backpropagation**, differentiation is a powerful concept in mathematics that can be used to optimize the weights. In mathematics, the derivative is the instantaneous rate of change of a function with respect to one of its variables. This is equivalent to finding the slope of the tangent line to the function at a point. Thus, we can define an “error signal”  $\delta_k$  at the output node that will be back-propagated toward the input. The error signal is calculated by Equation (2-8). In case of multi-layer, we can continue back-propagation the error signal toward the input by using Equation (2-9),  $\delta_k$  is passed through the output layer weights, summing overall output node, and passing the result through gradient of the activation function at hidden layer  $g'_j(z_j)$  as showed in Figure 2-7.

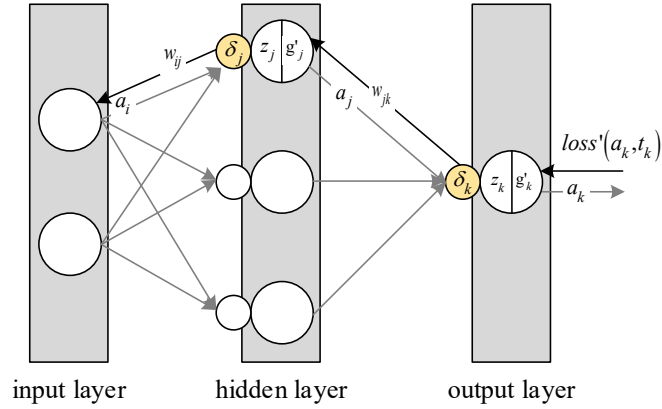


Figure 2-7 Back-propagate error signals

$$\delta_k = g'_k(z_k) loss'(a_k, t_k) \quad (2-8)$$

$$\delta_j = g'_j(z_j) \sum_k \delta_k w_{jk} \quad (2-9)$$

Then, gradients of the error function with respect to the model parameters at each layer  $l$  using the forward signal, and backward error signal is thus obtained by Equation (2-10).

$$\frac{\partial(loss)}{\partial w_{l-1,l}} = a_{l-1} \delta_l \quad (2-10)$$

In term of bias parameter, it turns out with the same gradients rule. Thus the bias gradients for layer  $l$  are simply showed in Equation (2-11)

$$\frac{\partial(\text{loss})}{\partial b_l} = \delta_l \quad (2-10)$$

Finally, the weights and biases for given layer are updated according to Equation (2-11) and Equation (2-12), respectively. The amount that the weights and biases are updated during training is referred as “learning rate” ( $\eta$ ). The learning rate in the training process has a small positive value, often in the range between 0.0 and 1.0.

$$w_{l-1,l} = w_{l-1,l} - \eta(a_{l-1}\delta_l) \quad (2-11)$$

$$b_l = b_l - \eta\delta_l \quad (2-12)$$

## 2.2.4 Convolutional Neural Networks

Convolutional neural networks (CNNs) is a subclass of NNs which have one or more convolutional layer for receiving and processing the multiple units input before passing it through fully connected Multi-Layer Perceptron (MLP) as showed in Figure 2-8.

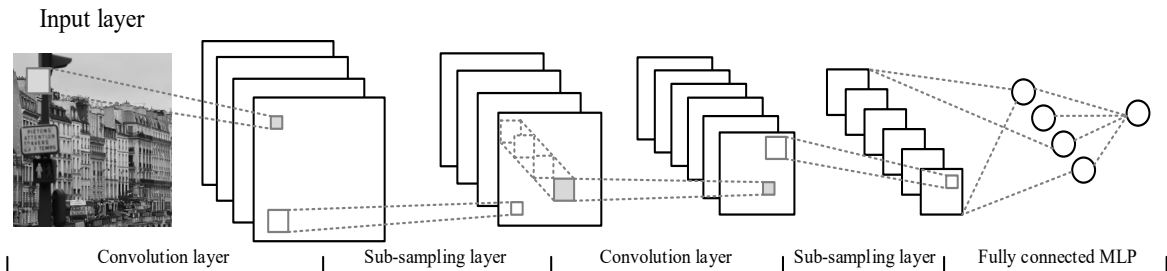


Figure 2-8 Convolution neural networks model

A simple convolutional neural networks is a sequence of layers, which consist of three main types of layers: Convolution layer, Sub-sampling layer or pooling layer, and Fully-connected layer (exactly as seen in regular Neural Networks). For the easiest way to explain the roll of each layer of CNNs, is to start simply with the example of image classification model for CIFAR-10

dataset. This dataset consists of many color images (sizes: 32x32) in 10 classes. The overview of architecture and each layers are described into more details below:

**Input layer** will hold pixel values of the image (width:32, height:32, 3 color channel)

**Convolutional layer** will compute the output of neurons that are connected to local regions in the input, each computing a dot product between their weights and a small region they are connected to in the input volume. This may result in volume such as [32x32x10] if we decided to use 10 filters.

**Sub-sampling layer (pooling)** will perform down-sample the output of a convolution layer along both the spatial dimensions of height and width. There are multiple techniques around pooling. Some of the most common pooling techniques are max pooling and average pooling. Figure 2-9 pooling process is demonstrated.

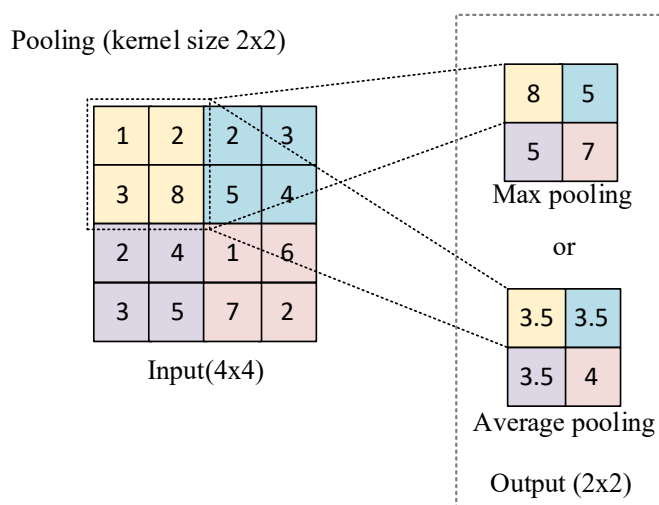


Figure 2-9 Pooling process

For this example, a 2 x 2 pooling operation on top of 10 feature maps will produce an output tensor of size [16 x 16 x 10].

**Fully-connected layer** will compute the class scores, resulting in volume of size [1x1x10], where each of the 10 numbers correspond to a class score, such as among the 10 categories of CIFAR-10. As with an ordinary Neural Networks and as the name implies, each neuron in this layer will be connected to all the numbers in the previous volume.

CNNs takes advantage of local spatial coherence in the input (often images) [12], which allow them to have fewer weights as some parameters are shared. This process, taking the form of convolutions, makes them especially well suit to extract relevant information at a low computational cost. We could say that CNNs outperform NNs on conventional image recognition tasks and many other tasks. Thus, CNNs have become the most popular tool in image recognition system [13], such, object recognition and face recognition.

## **2.3 Genetic Algorithm**

### **2.3.1 Introduction**

Genetic Algorithms are heuristic search approaches that are applicable to a wide range of optimization problems. This flexibility makes them attractive for many optimization problems in practice. Evolution is the basis of Genetic Algorithms. The current variety and success of species is a good reason for believing in the power of evolution. Species are able to adapt to their environment. They have developed to complex structures that allow the survival in different kinds of environments. Mating and getting offspring to evolve belong to the main principles of the success of evolution. These are good reasons for adapting evolutionary principles to solving optimization problems.

In this section, the foundations of Genetic Algorithms are presented. Starting with an introduction to the basic Genetic Algorithm with populations. Then, introduce the most important genetic operators step by step, which are crossover, mutation, and selection. Further, a genotype-phenotype mapping and common termination conditions will be explained.

### **2.3.2 Basic genetic algorithm**

The classic Genetic Algorithm is based on a set of candidate solutions that represent a solution to the optimization problem. A solution is a potential candidate for an optimum of the optimization problem. Its representation plays an important role, as the representation determines

the choice of the genetic operators. Representations are usually lists of values and are more generally based on sets of symbols. If they are continuous, they are called vectors, if they consist of bits, they are called bit strings. In case of combinatorial problems the solutions often consist of symbols that appear in a list. An example is the representation of a tour in case of the traveling salesman problem. Genetic operators produce new solutions in the chosen representation and allow the walk in solution space. The coding of the solution as representation, which is subject to the evolutionary process, is called genotype or chromosome. Algorithm 1 shows the pseudocode of the basic Genetic Algorithm, which can serve as blueprint for many related approaches. At the beginning, a set of solutions, which is denoted as population, is initialized. This initialization is recommended to randomly cover the whole solution space or to model and incorporate expert knowledge. The representation determines the initialization process. For bit string representations a random combination of zeros and ones is reasonable, for example, the initial random chromosome 1001001001 as a typical bit string of length 10. The main generational loop of the Genetic Algorithm generates new offspring candidate solutions with crossover and mutation until the population is complete.

Table 2-1 Basic Genetic Algorithm

<b>Basic Genetic Algorithm</b>
1: initialize population
2: <b>repeat</b>
3: <b>repeat</b>
4:   crossover
5:   mutation
6:   phenotype mapping
7:   fitness computation
8: <b>until</b> population complete
9: selection of parental population
10: <b>until</b> termination condition

### 2.3.3 Crossover

Crossover is an operator that allows the combination of the genetic material of two or more solutions. In nature, most species have two parents. Some exceptions do not know different sexes

and therefore only have one parent. In Genetic Algorithms, the crossover operators can be extended to more than two parents. The first step in nature is the selection of a potential mate partner. Many species spend a lot of resources on selection processes, but also on the choice of a potential partner and on strategies to attract partners. In particular, males spend many resources on impressing females. After the selection of a partner, pairing is the next natural step. From a biological perspective, two partners of the same species combine their genetic material and inherit it to their offspring.

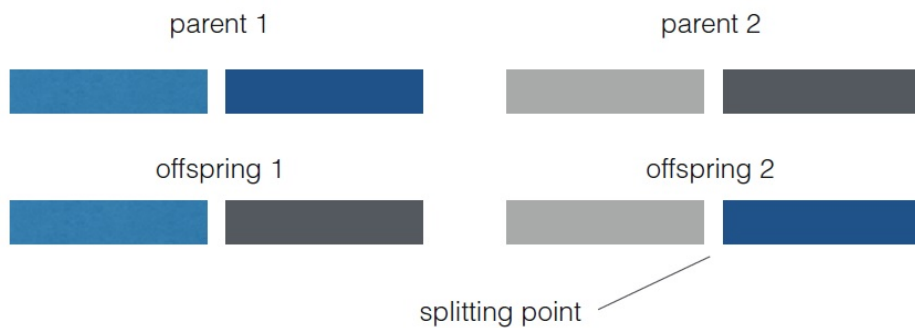


Figure 2-10 Illustration of one-point crossover and re-assembles to get two novel solutions

Crossover operators in Genetic Algorithms implement a mechanism that mixes the genetic material of the parents. A famous one for bit string representation is n-point crossover. It splits up two solution at n positions and alternately assembles them to a new one (Figure 2-10). For example, if 0010110010 is the first parent and 1111010111 is the second one, one-point crossover would randomly choose a position, let assume 4, and generate the two offspring candidate solutions 0010-010111 and 1111-110010. The motivation for such an operator is that both strings might represent successful parts of solutions that when combined even outperform their parents. This operator can easily be extended to more points, where the solutions are split up and reassembled alternately.

For continuous representations, the crossover operators are oriented to numerical operations. Arithmetic crossover, also known as intermediate crossover, computes the arithmetic mean of all parental solutions component-wise. For example, for the two parents (1, 4, 2) and (3,

2, 3) the offspring solution is (2, 3, 2.5). This crossover operator can be extended to more than two parents. Dominant crossover successively chooses each component from one of the parental solutions. Uniform crossover by uses a fix mixing ratio like 0.5 to randomly choose a bit from either of the parents. The question comes up, which of the parental solutions take part in the generation of new solutions. Many Genetic Algorithms simplify this step and randomly choose the parents for the crossover operation with uniform distribution.

### **2.3.4 Mutation**

The second protagonist in Genetic Algorithms is mutation. Mutation operators change a solution by disturbing them. Mutation is based on random changes. The strength of this disturbance is called mutation rate. In continuous solution spaces, the mutation rate is also known as step size.

There are three main requirements for mutation operators. The first condition is reachability. Each point in solution space must be reachable from an arbitrary point in solution space. An example that may complicate the fulfilment of this condition is the existence of constraints that shrink the whole solution space to a feasible subset. There must be a minimum chance to reach every part of the solution space. Otherwise, the chance is not positive that the optimum can be found. Not every mutation operator can guarantee this condition, for example, decoder approaches have difficulties covering the whole solution space.

The second good design principle of mutation operators is unbiasedness. The mutation operator should not induce a drift of the search to a particular direction, at least in unconstrained solution spaces without plateaus. In case of constrained solution spaces, bias can be advantageous.

The third design principle for mutation operators is scalability. Each mutation operator should offer the degree of freedom that its strength is adaptable. This is usually possible for mutation operators that are based on a probability distribution. For example, for the Gaussian mutation that is based on the Gaussian distribution, the standard deviation can scale the randomly



drawn samples in the whole solution space. The implementation of the mutation operators depends on the employed representation. For bit strings, bit flip mutation is usually used. Bit flip mutation flips a zero bit to a one bit and vice versa with a defined probability, which plays the role of the mutation rate. It is usually chosen according to the length of the representation. If  $N$  is the length of the bit string, each bit is flipped with mutation rate  $1/N$ . If the representation is a list or string of arbitrary elements, mutation randomly chooses a replacement for each element. This mutation operator is known as random resetting. Let  $[5, 7, -3, 2]$  be the chromosome with integer values that come from the interval  $[-10, 10]$ , then random resetting decides for each component if it is replaced. If the component is replaced, it randomly chooses a new value from the interval. For example, the result can be  $[8, -2, -5, 6]$ .

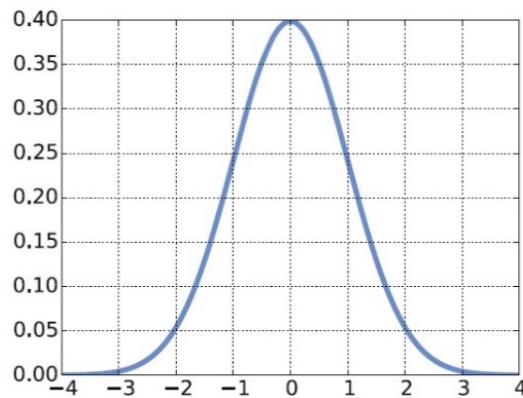


Figure 2-11 The Gaussian distribution

For continuous representations, Gaussian mutation is the most popular operator. Most processes in nature follow a Gaussian distribution, see Figure 2-11. This is a reasonable assumption for the distribution of successful solutions.

A vector of Gaussian noise is added to a continuous solution vector. If  $\mathbf{x}$  is the offspring solution that has been generated with crossover, is the Gaussian mutation with  $N(0, 1)$  as notation for a vector of Gaussian-based noise. Variable  $\sigma$  is the mutation rate that scales the strengths of the noise added. The Gaussian distribution is maximal at the origin. Hence, with the highest probability, the solution is not changed or only slightly. The Gaussian mutation is an excellent

example for a mutation operator that fulfills all mentioned conditions. With  $\sigma$  it is arbitrarily scalable. Moreover, with a scalable  $\sigma$ , all regions in continuous solution spaces will be reachable. Due to the symmetry of the Gaussian distribution, it does not prefer any direction and is hence drift less.

### **2.3.5 Genotype-phenotype mapping**

After crossover and mutation, the new offspring population has to be evaluated. Each candidate solution has to be evaluated with regard to its ability to solve the optimization problem. Depending on the representation a mapping of the chromosome, the genotype, to the actual solution, which is denoted as phenotype, is necessary. This genotype-phenotype mapping should avoid introducing a bias. For example, a biased mapping could map the majority of the genotype space to a small set of phenotypes. The genotype-phenotype mapping is not always required. For example, in continuous optimization, the genotype is the solution itself. But many other evolutionary modelling processes require this mapping.

### **2.3.6 Fitness**

In the fitness computation step, the phenotype of a solution is evaluated on a fitness function. The fitness function measures the quality of the solutions the Genetic Algorithm has generated. The design of the fitness function is part of the modeling process of the whole optimization approach. The practitioner can have an influence on design choices of the fitness function and thus guide the search. In case of multiple objectives that have to be optimized at the same time, the fitness function values of each single objective can be aggregated, for example by computing the weighted sum.

Most approaches aim at minimizing the number of fitness function calls. The performance of a Genetic Algorithm in solving a problem is usually measured in terms of the number of required fitness function evaluations until the optimum is found or approximated with a desired accuracy.

### 2.3.7 Selection

To allow convergence towards optimal solutions, the best offspring solutions have to be selected to be parents in the new parental population. A surplus of offspring solutions is generated and the best are selected to achieve a progress towards the optimum. This selection process is based on the fitness values in the population.

Many selection algorithms are based on randomness. Roulette wheel also known as fitness proportional selection selects parental solutions randomly with uniform distribution. The probability for being selected depends on the fitness of a solution. For this sake, the relative fitness of solutions normalized with the sum of all fitness values in a population, usually by division. This fraction of fitness can be understood as probability for a solution of being selected. The advantage of fitness-proportional selection operators is that each solution has a positive probability of being selected.

In case of comma selection, good parents can be forgot. Also, the randomness of fitness proportional selection allows forgetting of the best solutions. Although this might sound contra-productive for the optimization process at first, forgetting may be a reasonable strategy to overcome local optima. Another famous selection operator is tournament selection, where a set of solutions is selected randomly and within this competition subset, the best solutions are finally selected as new parents. The second step can be implemented with fitness proportional selection as typical example. Tournament selection offers a positive probability for each solution to survive, even if it has worse fitness values than other solutions.

When using selection as mechanism to choose the parents of the new generation, it is called survival selection. The selection operator determines, which solutions survive and which solutions die. This perspective directly implements Darwin's principle of survival of the fittest. But the introduced selection operators can also be employed for mating selection that is part of the crossover operators. Mating selection is a strategy to decide, which parents take part in the

crossover process. It makes sense to consider other criteria for mating selection than for survival selection.

### **2.3.8 Termination**

The termination condition defines when the main evolutionary loop terminates. Often, the Genetic Algorithm runs for a predefined number of generations. This can be reasonable in various experimental settings. Time and cost of fitness function evaluations may restrict the length of the optimization process. A further useful termination condition is convergence of the optimization process. When approximating the optimum, the progress of fitness function improvements may decrease significantly. If no significant process is observed, the evolutionary process stops. For example, when approximating the optima of continuous optimization problems, the definition of stagnation as repeated fitness difference lower than  $10^{-8}$  in multiple successive generations is reasonable. Of course, stagnation can only mean that the search might have got stuck in local optima, hence missing the global one. Restart strategies are approaches that avoid getting stuck in the same local optima. If the Genetic Algorithm repeatedly approximates the same area in solution space although starting from different areas, the chance is high that the local optimum is a large attractor, and a better local optimum is unlikely to find. It can also be that this local optimum is the global one.

## **2.4 Laban movement analysis**

### **2.4.1 Introduction**

Rudolf Laban (1879-1958) made significant contributions to the study of movement, bringing together his experiences as a dancer architect, painter, scientist, philosopher, and educator. He observed the movement of people performing types of tasks: from dancers to factory workers, fencers to people performing cultural ceremonies mental patients to managers and company executives. His theories on movement, which were significantly extended and applied

by his students and colleagues have resulted in a rich vocabulary for describing and analyzing movement, leading to the development of Laban Movement Analysis [14]. LMA has evolved into comprehensive system that has been used in dance, drama, nonverbal research, psychology, anthropology, ergonomics physical therapy, and many other movement-related fields.

LMA has five major components: Body, Space, Shape, Effort, and Relationship Together these components constitute a textual and symbolic language for describing movement. Body deals with the parts of the body that are used and the initiation and sequencing of a motion. Space describes the locale, directions, and paths of a movement. Shape involves the changing forms that the body makes in space. Effort describes how the body concentrates its exertion while performing movements. Effort is often compared to dynamic musical terms such as legato, forte, dolce, etc., which give information on how a piece of music should be performed. Relationship describes modes of interaction with oneself, others, and the environment. Relationship examples include facings, contact, and group forms.

In the HRI field, there have been many researchers who have applied LMA in order to defining relation between human motion and emotion [15], [16], [17]. The key parameter often used are LMA-Effort and LMA-Shape, as the authors in [18], [19] showed that these variables can be used effectively to encode emotional states. With this observation in mind, this thesis we attempt to develop a recognition model between human behaviour and emotion based on key features of LMA, which are Shape and Effort. The definition and description for Effort, and Shape will be explained in section 2.4.2 and section 2.4.3, respectively.

#### **2.4.2 LMA-effort**

Effort comprises four motion factors: space, Weight, Time, Flow. Each motion factor is a continuum between two extremes: (1) *indulging* in the quality and (2) *fighting* against the quality, In LMA these extreme Effort Elements are seen as basic, “irreducible” qualities, meaning they are the smallest units needed describing an observed movement. The eight Effort Elements are

Indirect/Direct, Light/Strong, Sustained/Sudden, and Free/Bound. The eight Elements can be combined and sequenced or innumerable variations of phrasings and expressions. Table 2-1 illustrates motion factors, listing their opposing Effort Elements with textual descriptions and examples.

Table 2-2 Motion factors and Effort elements

Motion Factors	Effort Elements
<b>Space:</b> attention to the surroundings	<b>Indirect:</b> flexible, meandering, wandering, multi-focus Examples: waving away bugs, slashing through plant growth
	<b>Direct:</b> single focus, channeled, undeviating Examples: pointing to a particular spot, threading a needle
<b>Weight:</b> sense of the impact of one's movement	<b>Light:</b> buoyant, delicate, easily overcoming gravity, marked by decreasing pressure Examples: paint on a canvas, describing the movement of feather
	<b>Strong:</b> powerful, having an impact, increasing pressure into the movement Examples: punching, pushing heavy object, expressing a firmly held opinion
<b>Time:</b> lack or sense of urgency	<b>Sustained:</b> lingering, leisurely, indulging in time Examples: stretching to yawn, stroking a pet
	<b>Sudden:</b> hurried, urgent Examples: swatting a fly, grabbing a child from the path of danger
<b>Flow:</b> attitude towards bodily tension and control	<b>Free:</b> uncontrolled, abandoned, unable to stop in the course of the movement Examples: waving wildly, shaking off water
	<b>Bound:</b> controlled, restrained, able to stop Examples: moving in slow motion chi, carefully carrying a cup of hot liquid

### 2.4.3 LMA-shape

The Shape component involves three distinct qualities of change in the form of movement: Shape Flow, Directional Movement, and Shaping. A Shape Flow attitude primarily reflects the mover's concern with the changing relationship among body parts. These changes can be sensed as the increasing or decreasing volume of the body's form or a moving toward or away from the

center of body. Shape Flow can be seen from these two different perspectives. The first one emphasizes the torso which can be said to “Grow” or “Shrink”. A continuous breathing pattern reveals changes in Shape Flow as seen from the torso perspective. The other perspective emphasizes the limbs, which are said to be Opening or Closing with respect to the longitudinal axis. Shrinking from the cold or stretching to wake up would be characterized as having a Shape Flow quality

While Shape Flow is mainly concerned with sensing the body's shape changes within itself, Directional Movement describes the mover's intent to bridge the action to point in the environment. These movements be simple spoke-like or arc-like actions to reach a direction or object, such as a reach to shake a hand or touch an object or to move specific location. Shaping Movement depicts the changes in movement form that demonstrate a carving, or moulding attitude as the body interacts the environment. This form can be dictated by objects in space or simply created by the mover. An active adapting of the body shape in order to move through a crowd, or a gesture describing an elaborately carved sculpture might illustrate a Shaping mode.

Shape changes in movement can be described in terms of three dimensions: Horizontal, Vertical and Sagittal. Each one of these dimensions is in fact associated with one of the three main dimensions (Width, Length, and Depth) as well as one of the three planes (Horizontal, Vertical, and Sagittal) related to the human body. Changes in Shape in the Horizontal dimension occur mainly the side-open and side-across directions; as the movement becomes planar there would be more of a forward-backward component added to the primary side component. Changes in the Vertical dimension are manifested primarily in the upward-downward directions; the plane would add more sideward component to the up-down. Finally, changes in the sagittal dimension are more evident in the body's depth or the forward-backward direction, planar movement would add an upward-downward component

Table 2-3 Shaping Dimensions and affinities

Shaping Dimensions	Affinities
Horizontal	<b>Spreading:</b> affinity with Indirect Examples: opening arm to embrace, sprawling in a chair
	<b>Enclosing:</b> affinity with Direct Examples: clasping someone in a hug, huddling in a cold
Vertical	<b>Rising:</b> affinity with Light Examples: reaching for something in a high shelf
	<b>Sinking:</b> affinity with Strong Examples: stamping the floor with indignation
Sagittal	<b>Advancing:</b> affinity with Sustained Examples: reaching to shake hand
	<b>Retreating:</b> affinity with Sudden Examples avoiding a punch

## 2.5 Markov model

The Markov model is the most commonly used to model randomly changing system. The model is created based on the concept of Markov property that, the next state depends only upon the current state. With this concept allows reasoning and computation with the model that would be otherwise intractable. Thus, it is possible to create a predictive model for the randomly given data obeys to Markov property.

Generally, there are four common Markov models used in different situation: Markov chain, Hidden Markov Model, Markov decision process, and partially observable Markov decision process. In this section, the Markov chains and Hidden Markov Model will be explained in detail.

### 2.5.1 Markov Chains

A Markov model consists of a list of the possible states of that system, the possible transition paths between those states and the state transition probabilities of those transitions. For representing a basic Markov model, the nodes in the model represent certain states and arrows which denote the probabilities of movement between states or the probabilities of getting out of the states. In the basic model, Let  $X_n, n = 0,1,2,\dots$ , is a discrete time stochastic process with a



discrete state space  $S$ , if the state space could be assumed as  $\{1, 2, \dots, N\}$  or  $\{0, 1, \dots, N-1\}$  in the finite state, and either  $\{0, 1, \dots\}$  or  $\{1, 2, \dots\}$  in the countably infinite state. To understand the behaviour of a process, Equation (2-13) shows the finite dimensional distributions, which allows for the calculation of any path probability, for every  $n$ , and every finite sequence of states  $i_0, i_1, \dots, i_n \in S$ .

$$P\{X_0 = i_0, X_1 = i_1, \dots, X_n = i_n\} = P\{X_0 = i_0, X_1 \in S, X_2 \in S, \dots, X_n = i_n\} \quad (2-13)$$

Above Equation, that is a common choice for such structure is the assumption that the processes satisfies the Markov property, that is to mention, the probability of next state depends only on the current state. It is called a discrete time Markov chain for any process  $X_n, n \geq 0$ , satisfying the Markov property as expressed by Equation (2-14).

$$P\{X_n = i_n | X_0 = i_0, \dots, X_{n-1} = i_{n-1}\} = P\{X_n = i_n | X_{n-1} = i_{n-1}\} \quad (2-14)$$

Furthermore, the one step transition probability of a Markov chain from state  $i$  to state  $j$  can be denoted as  $a_{ij}(n)$  that forms by Equation (2-15).

$$a_{ij}(n) = P\{X_n = j | X_{n-1} = i\}, \quad 1 \leq i, j \leq N \quad (2-15)$$

For the initial probability distribution of the process by  $\pi$ , which shows in Equation (2-16).

$$\pi_j = P\{X_0 = j\}, \quad j \in S \quad (2-16)$$

Returning to Equation (2-13) and then the problem of computing probabilities has been converted to one of simple multiplication as shown in Equation (2-17).

$$\begin{aligned}
& P\{X_0 = i_0, \dots, X_n = i_n\} \\
&= P\{X_n = i_n \mid X_0 = i_0, \dots, X_{n-1} = i_{n-1}\} \cdot P\{X_0 = i_0, \dots, X_{n-1} = i_{n-1}\} \\
&= a_{i_{n-1}, i_n} \cdot P\{X_0 = i_0, \dots, X_{n-1} = i_{n-1}\} \\
&= a_{i_{n-1}, i_n} \cdot P\{X_{n-1} = i_{n-1} \mid X_0 = i_0, \dots, X_{n-2} = i_{n-2}\} \cdot P\{X_0 = i_0, \dots, X_{n-2} = i_{n-2}\} \quad (2-17) \\
&= a_{i_{n-1}, i_n} \cdot a_{i_{n-2}, i_{n-1}} \cdot P\{X_0 = i_0, \dots, X_{n-2} = i_{n-2}\} \\
&\vdots \\
&= \pi_0 \cdot a_{i_0, i_1} \cdot a_{i_1, i_2} \cdot \dots \cdot a_{i_{n-2}, i_{n-1}} \cdot a_{i_{n-1}, i_n}
\end{aligned}$$

The first order or one step transition probabilities are mostly expressed in matrix form as expressed in Equation (2-18), where the matrix  $A$  is the state transition probability matrix for a Markov chain with state space  $S = \{1, 2, 3, \dots, N\}$ , each  $a_{ij}$  representing the probability of moving from state  $i$  to state  $j$  in the  $N \times N$  matrix.

$$A = \{a_{ij}\} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \dots & a_{NN} \end{pmatrix} \quad (2-18)$$

A summary of state transition probability matrix, each Markov chain can be defined as the transition probabilities,  $a_{ij}(n)$  at step  $n$  arranged into the probability matrix according to the current states as rows and the future states as columns, and row sums to one as shown in Equation (2-19).

$$\sum_{j \in S} a_{ij} = 1 \quad (2-19)$$

### 2.3.1.1 Example of the simple Markov model

The above stochastic process can consider as a sample 3 state Markov model of the weather as illustrated in Figure 2-12, where  $S_1, S_2, S_3, \dots, S_N$  a set of states,  $N$  is the number of state (for simplicity  $N = 3$ ). The weather is assumed that once a day and observed as being one of the

following: Sunny (State 1), Rainy (State 2) and Cloudy (State 3), and the probabilities of tomorrow's weather based on today's weather is expressed in Table 2-3.

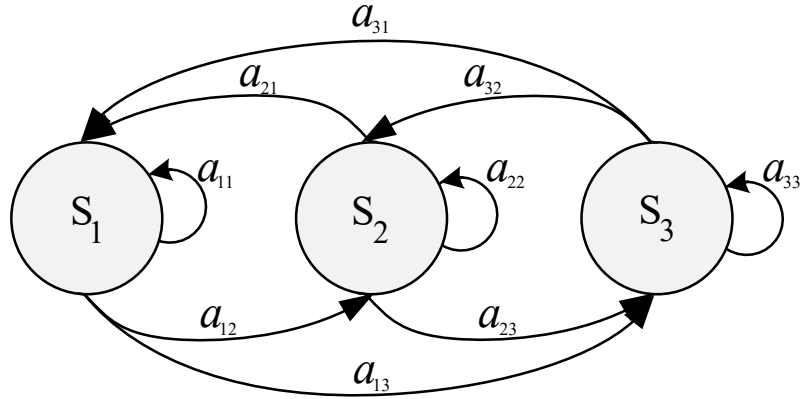


Figure 2-12 Markov model with three states.

Table 2-4 The probabilities for tomorrow's weather based on today's weather

Tomorrow's weather			Today's weather	
Sunny (S1)	Rainy (S2)	Cloudy (S3)		
0.8	0.1	0.1		Sunny (S1)
0.1	0.6	0.3		Rainy (S2)
0.2	0.3	0.5	Cloudy (S3)	

Let the weather on day 1 is Sunny (S1), ask the question: What is the probability that the weather for next 5 days will be “Sunny → Sunny → Rainy → Rainy → Cloudy ...”? For the formally, the sequences are arbitrary set as  $Q = \{S_1, S_1, S_1, S_2, S_2, S_3\}$  corresponding to the time  $(t)$ . The probability of  $Q$  can be expressed as below:

$$\begin{aligned}
 P(Q | Model) &= P[S_1, S_1, S_1, S_2, S_2, S_3 | Model] \\
 &= P[S_1] \cdot P[S_1 | S_1] \cdot P[S_1 | S_1] \\
 &\quad \cdot P[S_2 | S_1] \cdot P[S_2 | S_2] \cdot P[S_3 | S_2] \\
 &= \pi_1 \cdot a_{11} \cdot a_{11} \cdot a_{12} \cdot a_{22} \cdot a_{23} \\
 &= (1) \cdot (0.8) \cdot (0.8) \cdot (0.1) \cdot (0.6) \cdot (0.3) \\
 &= 1.152 \times 10^{-2}
 \end{aligned}$$

## 2.5.2 Hidden Markov Model

In simpler Markov models (Markov chain), the state is directly visible for observation, while in the Hidden Markov Model (HMM), the system state is only partially observable, which the invisible states are called “hidden states”. However, each hidden state has probability distribution over output states, which is observed (observable state).

The mechanism inside hidden layer is perform as a Markov process. In addition, for each of the possible states in hidden layer, there is a set of emission probabilities governing the distribution of the observed state at a particular time given the state of the hidden variable at that time.

An HMM is specified by the following components:

**A set of hidden states:**  $Q = \{q_1, q_2, \dots, q_N\}$

**State transition probability:** a transition probability matrix  $A$  as described in Equation (2-18), and Equation (2-19)

**An initial probability distribution:**  $\pi = \{\pi_1, \pi_2, \dots, \pi_N\}$ ,  $\sum_{n=1}^N \pi_n = 1$

**A sequence of observation:**  $O = \{o_1, o_2, \dots, o_T\}$

**Emission probabilities:**  $B = B(O_j|S_i) = B_{i,j}$  probability of an observation state  $j$  being generated from a hidden state  $i$ .

The general structure of HMM are illustrated in Figure 2-13, which generally consist of two main layers, hidden layer and observable layer.

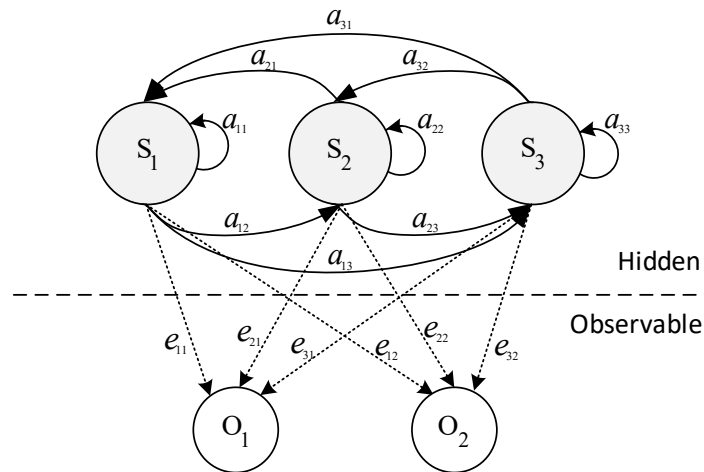


Figure 2-13 HMM example structure

An influential tutorial by Rabiner (1989), based on tutorials by Jack Ferguson in the 1960s, introduced the idea that hidden Markov models should be characterized by three fundamental problems:

**Problem 1 (Likelihood):** Given an HMM  $\lambda = (A, B)$  and an observation sequence  $O$ , determine the likelihood  $P(O|\lambda)$ .

**Problem 2 (Decoding):** Given an observation sequence  $O$  and an HMM  $\lambda = (A, B)$ , discover the most likely hidden state  $Q$ .

**Problem 3 (Learning):** Given an observation sequence  $O$  and the set of states in the HMM, learn the HMM parameters  $A$  and  $B$ .

In this thesis, we adopt HMM to decode the most likely hidden-state from the given observable sequence. The Viterbi algorithm is adopted to decrypt each emitter (state) in the hidden layer, which will be explained in detail in the next section.

### 2.5.3 Viterbi Algorithm

In this section, Viterbi algorithm will be described in more detail. Viterbi algorithm was proposed as a decoding algorithm, and has become standard terms for the application of dynamic programming algorithms to maximization problems involving probabilities. The Viterbi

algorithm provides an efficient way of finding the most likely state sequence, especially in the context of HMM [20], [21]. The explanation of how Viterbi algorithm work is explained below:

Given HMM with:

Observation space  $O = \{o_1, o_2, \dots, o_N\}$

Hidden state space  $S = \{s_1, s_2, \dots, s_K\}$

Initial probabilities  $\Pi = (\pi_1, \pi_2, \dots, \pi_K)$

A sequence of observations  $Y = (y_1, y_2, \dots, y_T)$

In order to simpler computation, the transition probability and emission probabilities are stored in the matrix form: where transition matrix  $A$  of size  $K \times K$  such that  $a_{ij}$  stores the transition probability of transiting from state  $s_i$  to state  $s_j$ ; emission matrix  $B$  of size  $K \times N$  such that  $B_{ij}$  stores the probability of observing  $o_j$  from state  $s_i$ . From the given parameter above, the most likely hidden state sequence  $X = (x_1, x_2, \dots, x_N)$  is thus obtained by computing with the following pseudocode.

Table 2-5 Viterbi algorithm

<b>Viterbi Algorithm</b>
1: Function Viterbi(O, S, $\Pi$ , Y, A, B): X
2: for each state $i \in \{1, 2, \dots, k\}$ do
3: $T_1[i, 1] \leftarrow \pi_i \cdot B_{iy_1}$
4: $T_2[i, 1] \leftarrow 0$
5: end for
6: for each observation $i = 2, 3, \dots, T$ do
7:     for each state $j \in \{1, 2, \dots, K\}$ do
8: $T_1[j, i] \leftarrow \max_k (T_1[k, i - 1] \cdot A_{kj} \cdot B_{jy_i})$
9: $T_2[j, i] \leftarrow \arg \max_k (T_1[k, i - 1] \cdot A_{kj} \cdot B_{jy_i})$
10:     end for
11: end for
12: $z_T \leftarrow \arg \max_k (T_1[k, T])$
13: $x_T \leftarrow s_{z_T}$
14: for $i \leftarrow T, T - 1, \dots, 2$ do
15: $z_{i-1} \leftarrow T_2[z_i, i]$
16: $x_{i-1} \leftarrow s_{z_{i-1}}$
17: end for
18: return X
19: end function

### 2.3.3.1 Example of the Viterbi algorithm for decoding hidden state

To simplify the explanation of how Viterbi algorithm decode state from HMM, as the model shown in Figure2-11, suppose you worked as a spy and was arrested in underground jails. However, it's a good thing that your enemy needs to know your classified information, so they keep you alive by giving you daily meals. Nevertheless, you need to escape from this place, and it is possible weather on cloudy day or the rainy day, because you can use the advantage of unclear vision to escape. Anyway, you cannot know the weather outside (Hidden state), the only piece of evidence you have is whether the person who comes into the room bringing your daily meal is carrying an umbrella or not (Observable state). The HMM parameters for this situation can be represented in Table. 2-6

Table 2-6 HMM parameters for weather prediction model

HMM	Model parameters
Initial prob.	Initial probabilities: $\Pi = \{0.333, 0.333, 0.333\}$
Hidden state sequences $X = (x_1, x_2, \dots, x_n)$ with $x_n \in S = \{s_1, s_2, \dots, s_K\}$	Weather state: $K=3$ ; $S = \{\text{Sunny, Rainy, Cloudy}\}$
Observable sequences $Y = (y_1, y_2, \dots, y_n)$ with $y_n \in O = \{o_1, o_2, \dots, o_N\}$	Umbrella state: $N=2$ ; $O = \{\text{carrying, not carrying}\}$
Transition prob.	$A = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.6 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}$ $a_{ij}$ representing the probability of moving from hidden state $i$ to hidden state $j$
Emission prob.	$B = \begin{bmatrix} 0.1 & 0.9 \\ 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix}$ $B_{i,j}$ or $B(O_j S_i)$ stores the probability of observing $o_j$ from state $s_i$

Assuming that the observed sequences of jailer behaviour is  $Y = \{O_1, O_2, O_2, O_1, O_2\}$ , where  $O_1$  and  $O_2$  stand for “Carrying” and “Not carrying” respectively. From the first day, the initial of all possible weather (Sunny, Rainy, Cloudy) is  $\Pi = \{0.333, 0.333, 0.333\}$ , and the transition of weather state ( $A$ ) is shown in the Table.2-2. Moreover, the weather is a condition for

the jailer in considering to carry an umbrella or not ( $B$ ). The question is: If you try to escape on the fifth day, is it will succeed? The computing process is demonstrated by following step:

**Step.1** Finding the maximum probabilities from both hidden states ( $h_{1,k}$ ) which might emitted the first sequence ( $t=1$ ), state “ $O_1$ ” by Equations (2-20), (2-21),

$$h_{1,k} = \pi_k \cdot B(y_1|k) \quad 2-20$$

$$x_1 = \text{argmax}(h_{1,k}) \quad 2-21$$

The results of each state from Equation (2-20) are  $h_{1,o_1} = (\pi_1) \cdot B(O_1|S_1) = 0.0333$ ,  $h_{2,o_1} = (\pi_2) \cdot B(O_1|S_2) = 0.2331$ , and  $h_{3,o_1} = (\pi_3) \cdot B(O_1|S_3) = 0.1332$ . So, we thus obtained the first hidden state by fed these results in to Equation (2-21). The result turned out that the observed state “ $O_1$ ” was emitted from hidden state “ $S_2$ ” (Rainy).

**Step.2** The previous possibilities result are used for calculating the next possible hidden state by Equations (2-22), (2-23),

$$h_{t,k} = \max_{x \in S} (h_{t-1,x} \cdot a_{xk} \cdot B(y_t|k)) \quad 2-22$$

$$x_t = \text{arg max}_{x \in S} (h_{t,k}) \quad 2-23$$

Thus, the second sequence of hidden state “ $O_2$ ” is thus obtained by computing with following parameters: possibilities of previous sequence ( $h_{t-1,x}$ ); transition probabilities of transitioning from previous mood state to next mood state ( $a_{xk}$ ); and emission probabilities of emitting state “ $O_2$ ” ( $B(O_2|k)$ ). The second step of calculation is demonstrated and shown in Figure2-14.



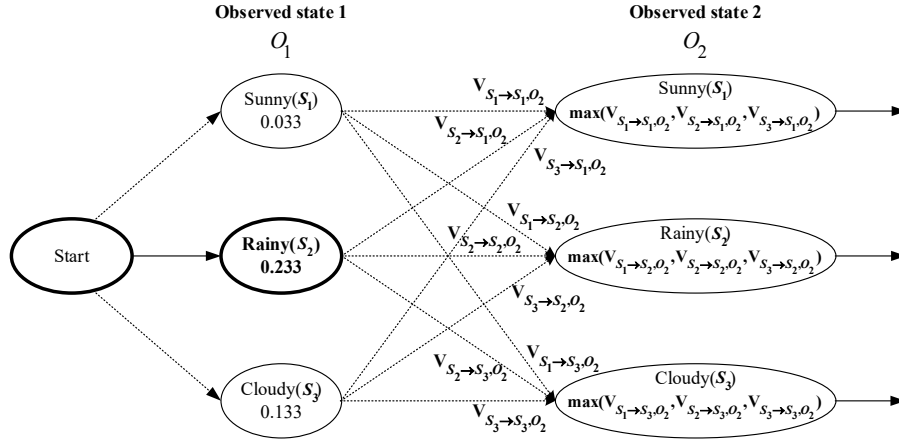


Figure 2-14 Graphical represent of Viterbi's calculation

The probabilities of each possible hidden state which emit observed state “ $O_2$ ”, is obtained as follows:

Sunny( $S_1$ ):

$$V_{S_1 \rightarrow S_1, O_2} = h_{1, O_1} \cdot a_{S_1 S_1} \cdot B(O_2 | S_1) = (0.0333)(0.8)(0.9) = 0.023976$$

$$V_{S_2 \rightarrow S_1, O_2} = h_{2, O_1} \cdot a_{S_2 S_1} \cdot B(O_2 | S_1) = (0.2331)(0.1)(0.9) = 0.020979$$

$$V_{S_3 \rightarrow S_1, O_2} = h_{3, O_1} \cdot a_{S_3 S_1} \cdot B(O_2 | S_1) = (0.1332)(0.2)(0.9) = 0.005994$$

Rainy( $S_2$ ):

$$V_{S_1 \rightarrow S_2, O_2} = h_{1, O_1} \cdot a_{S_1 S_2} \cdot B(O_2 | S_2) = (0.0333)(0.1)(0.2) = 0.000666$$

$$V_{S_2 \rightarrow S_2, O_2} = h_{2, O_1} \cdot a_{S_2 S_2} \cdot B(O_2 | S_2) = (0.2331)(0.6)(0.2) = 0.027972$$

$$V_{S_3 \rightarrow S_2, O_2} = h_{3, O_1} \cdot a_{S_3 S_2} \cdot B(O_2 | S_2) = (0.1332)(0.3)(0.2) = 0.007992$$

Cloudy( $S_3$ ):

$$V_{S_1 \rightarrow S_3, O_2} = h_{1, O_1} \cdot a_{S_1 S_3} \cdot B(O_2 | S_3) = (0.0333)(0.1)(0.6) = 0.001998$$

$$V_{S_2 \rightarrow S_3, O_2} = h_{2, O_1} \cdot a_{S_2 S_3} \cdot B(O_2 | S_3) = (0.2331)(0.3)(0.6) = 0.041958$$

$$V_{S_3 \rightarrow S_3, O_2} = h_{3, O_1} \cdot a_{S_3 S_3} \cdot B(O_2 | S_3) = (0.1332)(0.5)(0.6) = 0.039960$$

From the probabilities result above, the maximum probability  $V_{S_2 \rightarrow S_3, O_2} = 0.041958$  appear in the hidden state  $S_3$  or “Cloudy”. The result suggested that when the jailer did not carry umbrella on the second day of observation ( $O_2$ ), the weather outside is “Cloudy”.

**Step.3** repeatedly calculate in Step 2 through all of observation sequences. The sequences of jailer behaviour ( $O_1, O_2, O_2, O_1, O_2$ ) were emitted from these hidden state sequences ( $S_2, S_3, S_1, S_2, S_1$ ). So it is not save to escape on the fifth day because it is “Sunny day”

## 2.6 Summary

In this chapter, the fundamental theories: Neural networks for robotics system, Laban movement analysis theory, Markov model, and Motion control system, as described above, will be used for developing the human-robot emotional interactive system. The methodology of using mentioned theories will be discussed in Chapter 3. Thus endowing the robot with the emotional skill, will enhance the robot with a better interaction in a more human-oriented way.

## Chapter 3

### Methodology

As mentioned in Chapter 1, HRI is the system that work along human and should be able to seamlessly perform a task like human. The Seamless robot was designed with functionality comparable to the human upper body consisting of trunk, head and two bilateral arms. Upper-body expression of the Seamless robot has two elements, namely, head and arm. Head movement consists of 1) nodding motions and 2) side to side twisting. Arm movement is driven by the two shoulder joints, one twisting joint on the upper-arm, an elbow joint, a twisting joint in the lower-arm, wrist joint, and hand gripper as shown in Figure 3-1. The robot's arm and face are used in a model of emotional expression. A depth camera is used as a robot's sensory device to receive an input from human, both face and body joint movement. The detail of Seamless robot structure will be described then follow by the control scheme, the implementation of neural networks for solving robotic arm kinematic solution and overview of Seamless robot's software will be clearly explained in this chapter.

#### 3.1 Robot system configuration

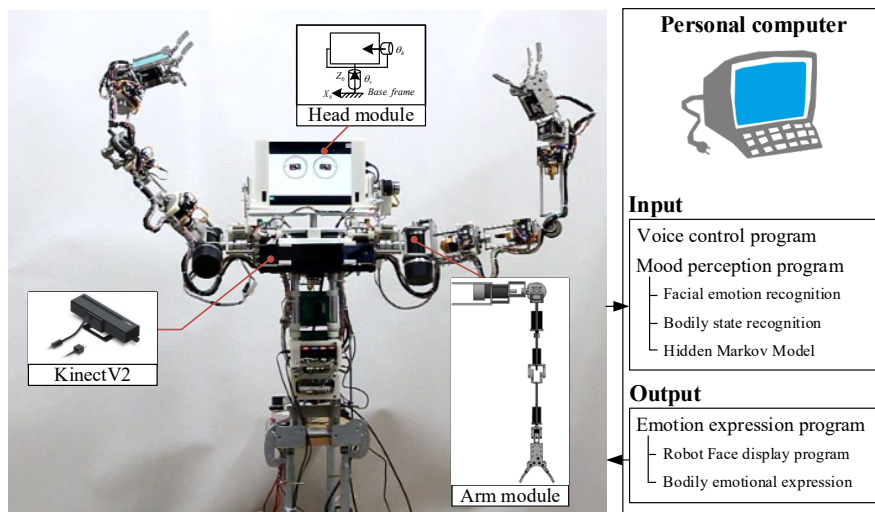


Figure 3-1 Overview of Seamless robot system

### 3.1.1 Hardware configuration

#### 3.1.1.1 Head module configuration

In this part, a monitor is used to display the eye emotions of Seamless robot as shown in Figure 3-2. The eye movement simulator is created by the Open Graphics Library (OpenGL) software in order to express the basic emotions of the robot. Two brushless direct current motors are installed to create two DOFs (roll and yaw) in the neck. The specification of actuators are listed in Table 3-1 and Table 3-2.

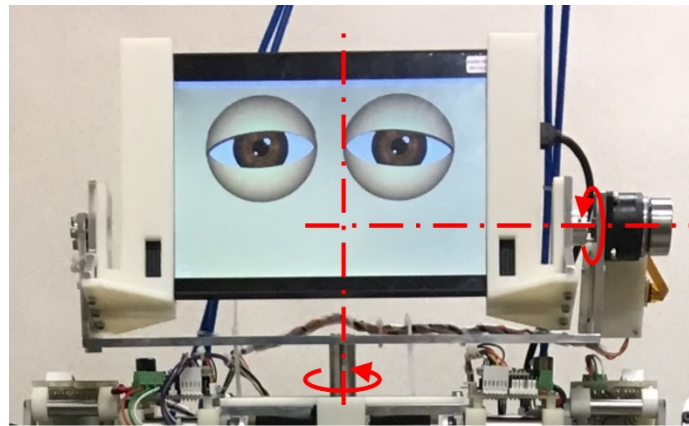


Figure 3-2 Configuration of head module

Table 3-1 Specification of neck roll joint motor



	<p><b>Motor:</b> Maxon EC 32 flat Ø32 mm, brushless, 15 Watt, with Hall sensors, Part No.: 267121</p> <ul style="list-style-type: none"> <li>• Outer diameter: 32 mm</li> <li>• Type power: 15 W</li> <li>• Nominal voltage: 24 V</li> <li>• No load speed: 4530 rpm</li> <li>• Nominal torque (max. continuous torque): 25.5 mNm</li> </ul>
	<p><b>Gear:</b> Maxon Spur Gearhead GS 38 A Ø38 mm, 0.1 - 0.6 Nm, Part No.: 110456</p> <ul style="list-style-type: none"> <li>• Outer diameter: 38 mm</li> <li>• Reduction: 100 : 1</li> <li>• Max. continuous torque: 0.2 Nm</li> </ul>

Table 3-2 Specification of neck yaw joint motor

	<p><b>Motor:</b> Maxon EC 32 flat Ø32 mm, brushless, 15 Watt, with Hall sensors, Part No.: 267121</p> <ul style="list-style-type: none"> <li>• Outer diameter: 32 mm</li> <li>• Type power: 15 W</li> <li>• Nominal voltage: 24 V</li> <li>• No load speed: 4530 rpm</li> <li>• Nominal torque (max. continuous torque): 25.5 mNm</li> </ul>
	<p><b>Gear:</b> Maxon Spur Gearhead GS 38 A Ø38 mm, 0.1 - 0.6 Nm, Part No.: 110454</p> <ul style="list-style-type: none"> <li>• Outer diameter: 38 mm</li> <li>• Reduction: 30 : 1</li> <li>• Max. continuous torque: 0.2 Nm</li> </ul>

### 3.1.1.2 Arm module configurations

In seamless robot, the arm has been designed to have a structure like human arm. The total length of each arm part is 695 mm. Figure 3-3 shows the structure of robot arm. Each arm has 7 levels of flexibility: the shoulder (joint 1, joint 2), the elbow (joint 3, joint 4), the wrist (joint 5, joint 6), and The hand part, which has 3 fingers. The specification of shoulder, elbow, wrist, and hand actuators are listed in Table 3-3, Table 3-4, Table 3-5, and Table 3-4 respectively.

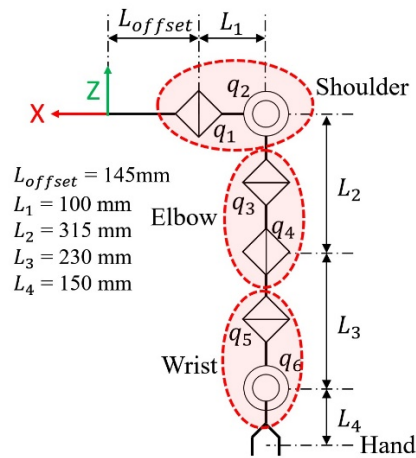


Figure 3-3 Structure of robot arm

Table 3-3 Specification of shoulder joint motor





$q_1$		<p><b>Motor:</b> Maxon EC 60 flat Ø60 mm, brushless, 100 Watt, with cover, Part No.: 408057</p> <ul style="list-style-type: none"> <li>• Outer diameter: 60 mm</li> <li>• Type power: 100 W</li> <li>• Nominal voltage: 24 V</li> <li>• No load speed: 4250 rpm</li> <li>• Nominal torque (max. continuous torque): 227 mNm</li> </ul>
		<p><b>Gear:</b> Maxon Planetary Gearhead GP 52 C Ø52 mm, 4 - 30 Nm, Ceramic Version, Part No.: 223095</p> <ul style="list-style-type: none"> <li>• Outer diameter: 52 mm</li> <li>• Reduction: 113: 1</li> <li>• Max. continuous torque: 30 Nm</li> </ul>
$q_2$		<p><b>Motor:</b> Maxon EC 60 flat Ø60 mm, brushless, 100 Watt, with cover, Part No.: 408057</p> <ul style="list-style-type: none"> <li>• Outer diameter: 60 mm</li> <li>• Type power: 100 W</li> <li>• Nominal voltage: 24 V</li> <li>• No load speed: 4250 rpm</li> <li>• Nominal torque (max. continuous torque): 227 mNm</li> </ul>
		<p><b>Gear:</b> Maxon Planetary Gearhead GP 52 C Ø52 mm, 4 - 30 Nm, Ceramic Version, Part No.: 223095</p> <ul style="list-style-type: none"> <li>• Outer diameter: 52 mm</li> <li>• Reduction: 113: 1</li> <li>• Max. continuous torque: 30 Nm</li> </ul>

Table 3-4 Specification of elbow joint motor





q <sub>3</sub>		<p><b>Motor:</b> Maxon EC 45 flat Ø42.9 mm, brushless, 30 Watt, with Hall sensors, Part No.: 339281</p> <ul style="list-style-type: none"> <li>• Outer diameter: 45 mm</li> <li>• Type power: 30 W</li> <li>• Nominal voltage: 24 V</li> <li>• No load speed: 4360 rpm</li> <li>• Nominal torque (max. continuous torque): 54.7 mNm</li> </ul>
		<p><b>Gear:</b> Maxon Planetary Gearhead GP 42 C Ø42 mm, 3 - 15 Nm, Ceramic Version, Part No.: 260553</p> <ul style="list-style-type: none"> <li>• Outer diameter: 42 mm</li> <li>• Reduction: 216: 1</li> <li>• Max. continuous torque: 15 Nm</li> </ul>
q <sub>4</sub>		<p><b>Motor:</b> Maxon EC 32 flat Ø32 mm, brushless, 15 Watt, with Hall sensors, Part No.: 267121</p> <ul style="list-style-type: none"> <li>• Outer diameter: 32 mm</li> <li>• Type power: 15 W</li> <li>• Nominal voltage: 24 V</li> <li>• No load speed: 4530 rpm</li> <li>• Nominal torque (max. continuous torque): 25.5 mNm</li> </ul>
		<p><b>Gear:</b> Maxon Planetary Gearhead GP 32 C Ø32 mm, 1.0 - 6.0 Nm, Ceramic Version, Part No.: 166949</p> <ul style="list-style-type: none"> <li>• Outer diameter: 32 mm</li> <li>• Reduction: 246 : 1</li> <li>• Max. continuous torque: 6 Nm</li> </ul>

Table 3-5 Specification of wrist joint motor

q <sub>5</sub>		<p><b>Motor:</b> Maxon EC 32 flat Ø32 mm, brushless, 15 Watt, with Hall sensors, Part No.: 267121</p> <ul style="list-style-type: none"> <li>• Outer diameter: 32 mm</li> <li>• Type power: 15 W</li> <li>• Nominal voltage: 24 V</li> <li>• No load speed: 4530 rpm</li> <li>• Nominal torque (max. continuous torque): 25.5 mNm</li> </ul>
		<p><b>Gear:</b> Maxon Planetary Gearhead GP 32 C Ø32 mm, 1.0 - 6.0 Nm, Ceramic Version, Part No.: 166948</p> <ul style="list-style-type: none"> <li>• Outer diameter: 32 mm</li> <li>• Reduction: 190 : 1</li> <li>• Max. continuous torque: 6 Nm</li> </ul>
q <sub>6</sub>		<p><b>Motor:</b> Maxon EC 32 flat Ø32 mm, brushless, 15 Watt, with Hall sensors, Part No.: 267121</p> <ul style="list-style-type: none"> <li>• Outer diameter: 32 mm</li> <li>• Type power: 15 W</li> <li>• Nominal voltage: 24 V</li> <li>• No load speed: 4530 rpm</li> <li>• Nominal torque (max. continuous torque): 25.5 mNm</li> </ul>
		<p><b>Gear:</b> Maxon Spur Gearhead GS 38 A Ø38 mm, 0.1 - 0.6 Nm, Part No.: 110455</p> <ul style="list-style-type: none"> <li>• Outer diameter: 38 mm</li> <li>• Reduction: 60 : 1</li> <li>• Max. continuous torque: 0.2 Nm</li> </ul>





Figure 3-4 Dynamixel RX-24F

Table 3-6 Specification of hand motor

Specification	Model
	RX-24F
Manufacturer	Dynamixel
Weight	67g
Dimension	35.6mm x 50.6mm x 35.5mm
Resolution	0.29°
Gear Reduction Ration	193: 1
Stall Torque	2.6N.m (at 12V, 2.4A)
No load speed	126rpm (at 12.V)
Running Degree	0° ~ 300°
Running Temperature	-5°C ~ +80°C
Voltage	9V~12V
Link (Physical)	RS485 Multi Drop Bus
ID actuator	254 ID (0~253)
Communication Speed	7843bps ~ 1 Mbps
Types of Feedback	Position, Temperature, Load, Input Voltage, etc.

### 3.1.1.3 Control scheme

For shoulder, elbow, and wrist the motors that were installed do not come with a position sensor, instead, the manufacturer offers a hall sensor installed inside the motor. And because these

motor are BLDC motor, driving these motor require a specific motor driver. The motor driver that manufacturer provides is speed control motor driver, details are described in Table 3-7. However, the position control which is crucial in controlling the robot movement is not available from the device manufacturer provides.

To create a position controller, a microcontroller STM32F4 Discovery (Figure3-6) is installed. This microcontroller is being used for process the signal from hall sensor into incremental position, receive angle command from computer, and control the angle with PID controller. Connection of the control system is shown in Figure3-7.

Table 3-7 Specification of motor driver

Specification	Model	
	ESCON Module 50/4 EC-S, 4-Q Servocontroller	ESCON 36/3 EC, 4-Q Servocontroller
Manufacturer	Maxon motor	Maxon motor
Power	200 W	100W
Speed control	Open / Close loop control	Open / Close loop control
Operating voltage	10V-50V	10V-36V
Maximum output current	12A	9A
Input	Enable, Direction, PWM set value	Enable, Direction, PWM set value
Operating temperature	-30 °C - 45 °C	-30 °C - 45 °C
Weight	11 g	36 g
Dimension	43.2 mm x 31.8 mm x 12.7 mm	55 mm x 40 mm x 19.8 mm
Implement with	$q_1, q_2$	$q_3, q_4, q_5, q_6$

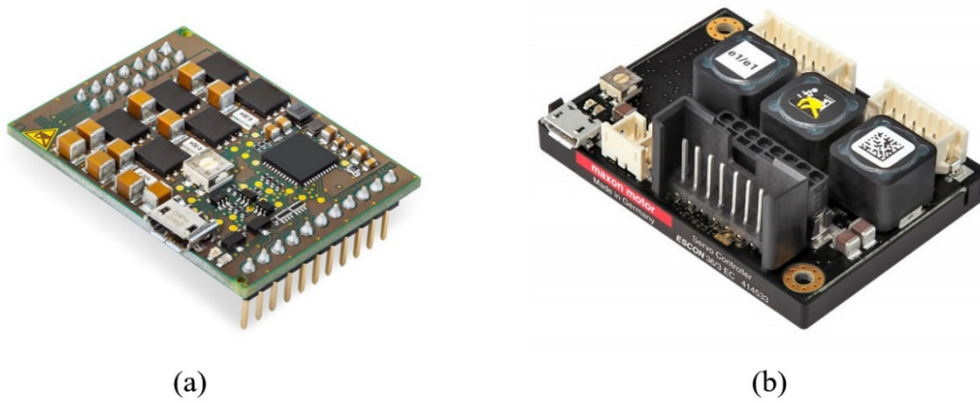


Figure 3-5 Motor driver  
 (a) ESCON Module 50/4 EC-S, (b) ESCON 36/3 EC

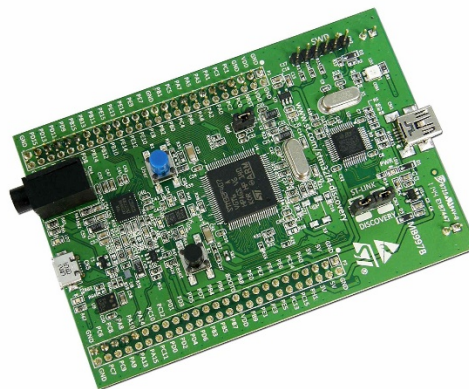


Figure 3-6 STM32F4 Discovery

Table 3-8 Specification of STM32F4 Discovery

Specification	Model
	STM32F4 Discovery
Processor	32-bit ARM® Cortex®-M4 with FPU core
Flash memory	1 Mbyte
RAM	192 Kbyte
Supply voltage	5V
Package	LQFP100

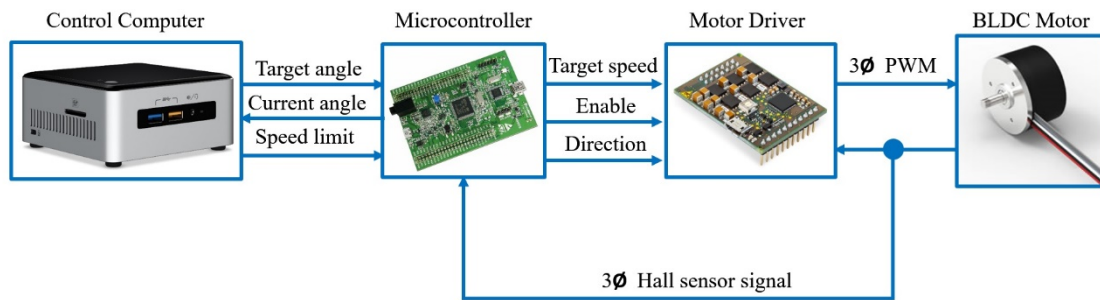


Figure 3-7 Diagram of control system

### 3.1.2 RGB-Depth camera (KinectV2)

KinectV2 is a line of motion sensing input device produced by Microsoft. Based around a webcam-style add-on peripheral, it enabled users to control and interact with their computer without the need for a control device, through a natural user interface using gestures and microphone array also provided for supporting voice commands. Therefore, we decided to use this device as perceptual equipment for install to a Seamless robot.



Figure 3-8 KinectV2

#### 3.1.4.1 Sample inputs obtained from KinectV2

Kinect V2 provides several raw input data, such, RGB frame image, Depth frame data, infrared image, and also microphone for capturing audio data. Moreover, Microsoft also provided brilliant software development kit (SDK) for Kinect V2, which enables developer to access features such as skeleton tracking, hand interactions, and voice recognition. The sample inputs are illustrated by Figure 3-9 and Figure 3-10.

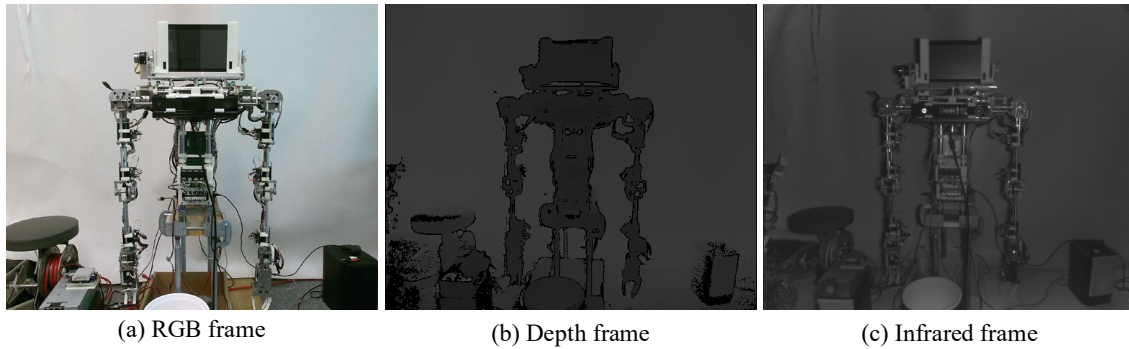


Figure 3-9 Multi-frame data (RGB frame, Depth frame, and Infrared frame)

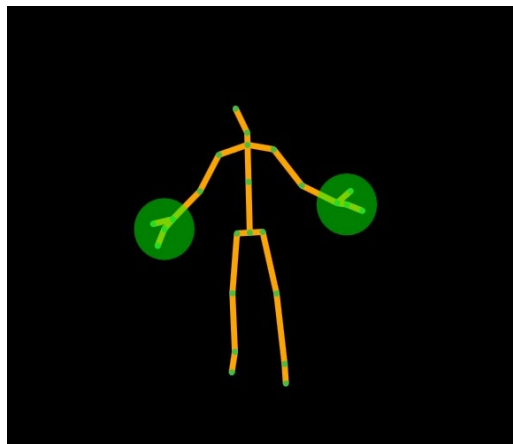


Figure 3-10 Skeleton tracking.

The skeleton tracking module provides body joints position, and it will be used as the raw data for studying human behaviour in emotional expression. The bodily emotional expression, consist of data collecting process, and feature extraction will be described detail genuinely in next section.

### 3.2 Study of emotional expression

This section talked about the definition of words feeling, which clearly explain, how different between emotions and moods — followed with the study procedure, the process for collecting data, how humans express their feelings through their physical movement. A method for extracting state of bodily expression based on Laban movement theory will be described as well in this section.

### 3.2.1 Emotions, Feelings, Moods, and its expression

Emotions, feelings and Moods are often used interchangeably, as a word for describing feelings. However, they really are not the same thing.

The emotions is the most basic level, so emotion is experienced. It is an automatic reaction. It occurs in a part of the brain that is not connected to cognitive reasoning which means that can be illogical, but it is going to be fast. There are several models of emotions. The basic emotions model [22] usually can be recognized, is the six emotions model. The six emotions model captures the range of emotions (**anger, disgust, fear, happiness sadness, and surprise**) that could be easily understood. These six basic emotions are also easily recognized on human facial expression.

Feelings, on the other hand, require cognitive awareness. So, we could think of feelings as emotions that are processed by thinking. Feelings tend to be long-lasting, less intense, and they occur when emotions are assigned meaning.

Mood is a state of feeling, and usually thought of as complex having emotions and multiple feelings together. If feelings are long-lasting and emotions are short lasting, so mood could be long lasting because feelings have long-lasting component. Additionally, moods have an effect on bodily expression. However, in the work reported in [23] which measured position, velocity and acceleration of certain body joint movements of the design team members for real time emotional state detection. According to the authors, the complexity of human motion is such that the collection of large amounts of joint data does not always lead to improved system performance, so the selection of which joints to track needs to be studied further.

### 3.2.2 Data collecting process

In our earlier work [24], we have studied about body emotional expression of human by asking twenty-six subjects (14 male and 12 female, 20 to 34 years old) were asked to exhibit bodily gestures to express their feelings both in positive and negative emotional states, within 3

seconds in each emotional scheme. A depth camera (Kinect V2) was used to capture the movements of upper body joints in humans. Upper-body joints data were obtained to compute the LMA-shape, triangular-shape, three points of upper-body limbs (head, left and right arm) are connected as illustrated in Figure 3-11

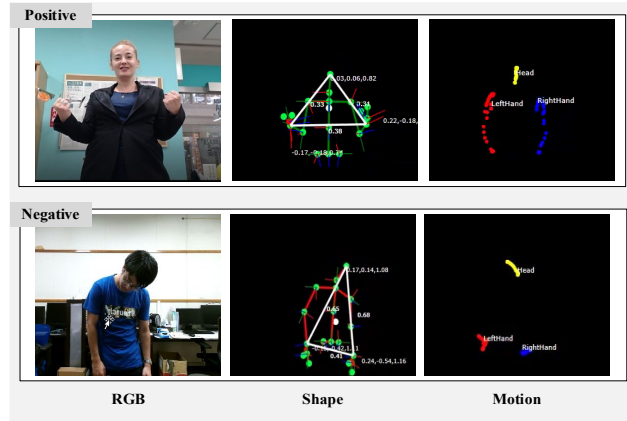


Figure 3-11 Emotional expression by human

As many researchers presented the use of LMA features for assessing human emotion as the work reported [25], which is useful for interpreting human motion. The complexity of human expression was simplified utilizing LMA features, LMA-shape and LMA-effort. LMA-shape applies to the connection between body parts and the way shapes change in space when the body moves. On the other hand, LMA-effort represents the quality of dynamic motion in actions.

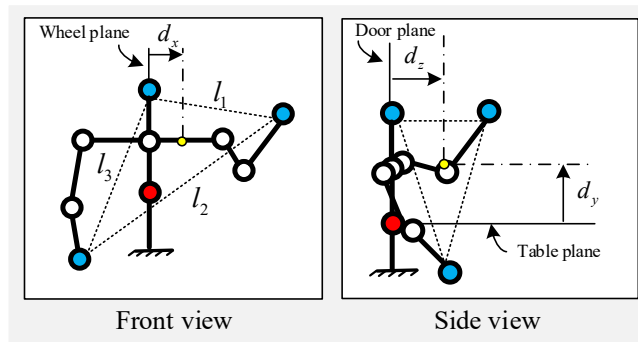


Figure 3-12 Triangle shape of upper body

In Figure 3-12 that presented the triangle features, distances of centroid position ( $d_x, d_y, d_z$ ), and the shape length ( $l_1, l_2, l_3$ ), which were used for defining body postures.

Distances of centroid position are used for describing of body-raised form, and a distance between both hands is used for discriminating between body-opened or body-closed form.

The study results of the shape feature of each emotional scheme are shown in Table 3-9. The analysed data showed that the participants in the experiment tended to raise their ( $d_y$ ) hand when in a positive mood higher than negative mood. Moreover, distances between both hands ( $l_2$ ) of these two emotional expressions have significantly different.

Table 3-9 Triangle parameters

	Parameter			
	$d_x(m)$	$d_y(m)$	$d_z(m)$	$l_2(m)$
<b>Positive</b>	0.003	0.316	0.088	0.623
<b>Negative</b>	-0.002	0.122	0.123	0.302

In the experiment described above, the participants were asked to express their feelings through bodily movement without any emotional stimulation. Arguably, the emotional expression data from this experiment is not a very good indicator of inner feeling. The authors in [26] proposed to infer expressions of emotion from motions observed during narrations. Data were collected from subjects, who were able to exhibit all forms of expression (i.e., facial, conversation and bodily movement) during the narration. The subjects were asked to use an annotation system and assign one emotion label to each utterance, thus effectively creating personalized acting scripts for themselves. In contrast, we wish to study the expressions of individuals experiencing more natural feeling. In the work reported in [27], remembering an event, a situation, or a person can evoke feelings, although feelings activated by a memory may not be felt as intensely as the actual experience. Note that recalling a poignant memory can trigger strong emotion. Based on this observation, we proposed a way to collect more reliable data, by asking subjects to recall past experiences, associated with both good and bad moods. The subjects were given brief verbal explanations of the relation between emotion and expression. Then, they were asked to recall past experiences, and to take turns sharing their experience with others.



The experiences recounted were classified as positive or negative, taking account of differences in perception between speaker and listener. If they disagreed, the story recounted was not included in the analysis, since the expression of the speaker might be misleading. While talking, the face image and body joints positions were recorded by a depth camera (KinectV2). To protect privacy sound was not recorded. The data obtained with this method is more reliable than that resulting from direct questioning. However, it is difficult to find the people who willing to share their life-experiences. For this reason, the analysed narrative data were obtained from a small sample consisting of 5 males and 5 females 20 to 32 years old. The data collection environment was arranged, as shown in Figure3-13.

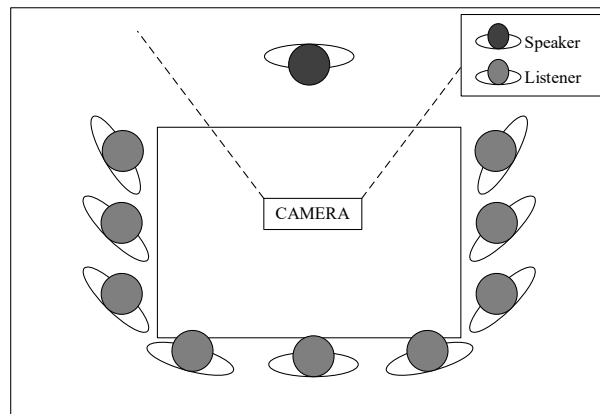


Figure 3-13 Data collecting desk

### 3.2.3 Body states extraction

The bodily expression data from both negative and positive mood scenes obtained from the 10 participants showed the diversity human movement. To reduce complexity, we adopted LMA-features to categorize human behaviours into simple states of expression.

#### 3.2.3.1 Posture state (Immobile)

In Figure3-14 (a) and Figure3-14 (b), the study samples of motion behaviour of the narration data are illustrated. The position of centroid motions showed that the distance in z-axis and y-axis have more significant information than the distance in the x-axis, consistent with the previous studied result. The red circle and dash-line are indicating a median position of data and

percentile of negative motions respectively. The bodily-expressions data from participant show that when they were sharing about a positive story, their centroids of motion widely spread out of the negative boundary of motion. In case of if a position in the y-axis is over the negative boundary, we defined it as raised posture.

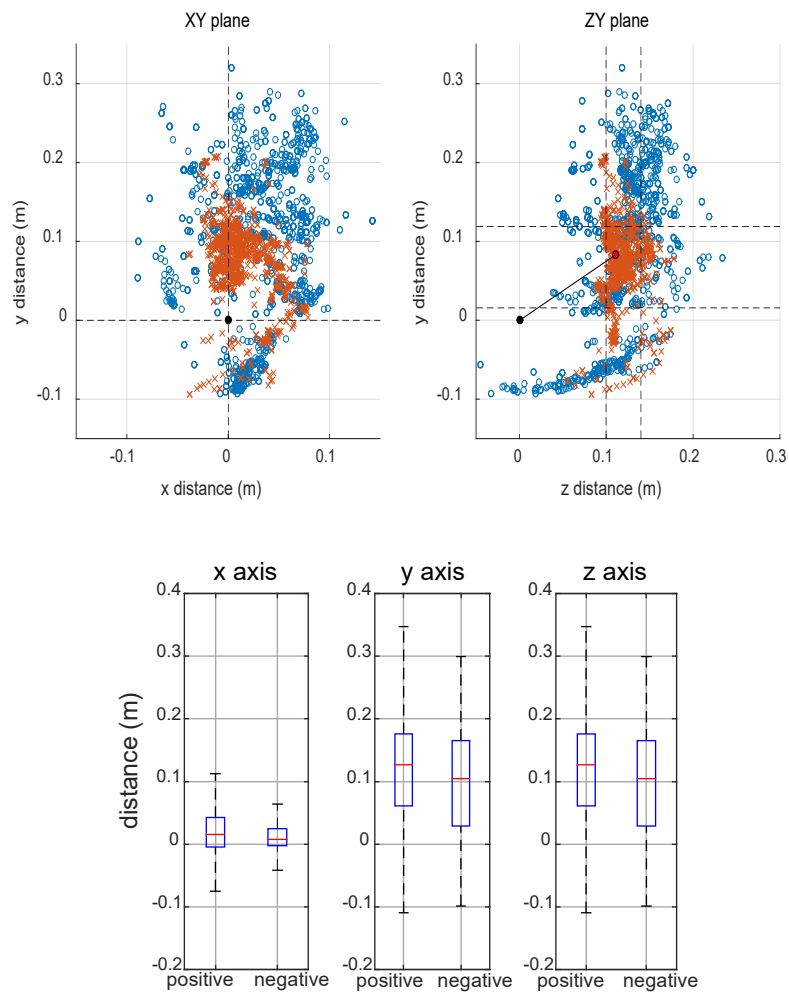


Figure 3-14 Emotional expression data  
 Top: Sample of centroid motion; Bottom: Boxplot of centroid position from studied data

Regarding the previous studied, left-hand to right-hand distance could be considered as supplementary to interpret other human posture. In Figure3-15, hand distance behaviour showed that, during the narrator sharing a positive story, distances between his or her hand are often over

their shoulder width (dash-line). In contrast, they often move both hands close to each other while sharing a negative story. The distance between both hands is used to describes posture, namely, opened or closed form. If the distance is more than shoulder width (dash line) of the narrator, it refers to open form and closed form in contrast

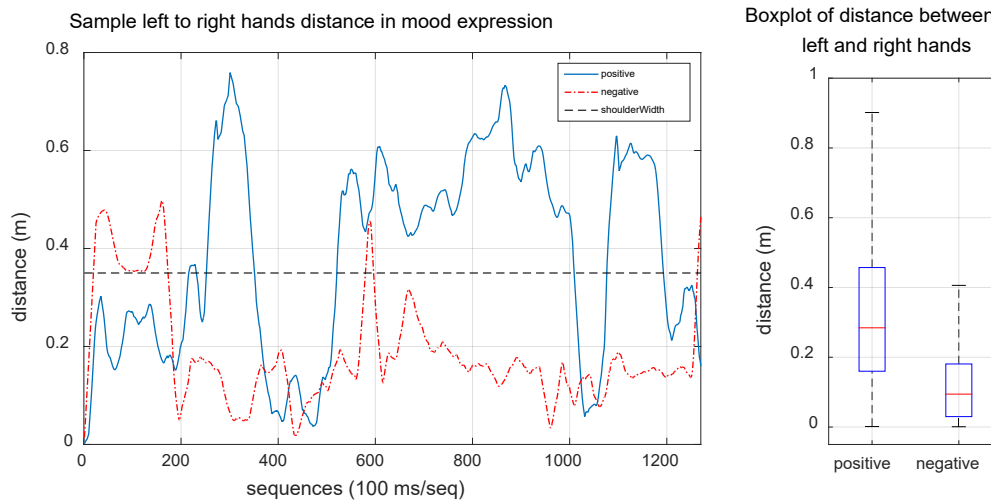


Figure 3-15 Left- to right-hands distance

The distance between both hands is used to describes posture, namely, opened or closed form. If the distance is more than shoulder width (dash line) of the narrator, it refers to open form and closed form in contrast..

### 3.2.3.2 Gesture states (Mobile)

While LMA-shape length describes body postures, LMA-effort describes characteristics of movement in actions, i.e., speed, and direction movement. The speed of hand movement is determined by distance changing from the previous hand position to the current position. During narration, when participants move their hands, the speed of both hand, left-hand and right-hand are essentially equivalent. Hence, gesture speed features were obtained by computing from both hands motion. In Figure3-16 quality of hand movement, both in the negative and positive scene are illustrated. Study result showed that during narrator sharing a positive story, they moved both hands with a higher average velocity than while sharing a negative story, as indicated with dash-line.

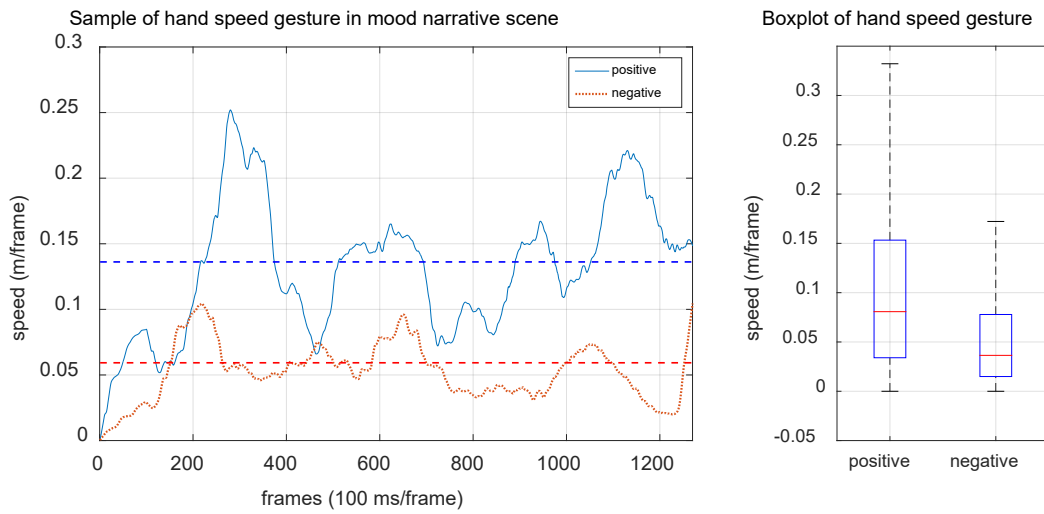


Figure 3-16 Hand gesture speed

Another parameter to verifying the quality of movement is considering from frequency of elbow angle movement due to its movement, has strongly affected to the direction changing of hand tip. The angular motion of both elbows was fed to compute the frequency response of movement. In Figure3-17, the hand movements of the narrator while telling a positive story also exhibit higher direction changing rate than a negative story.

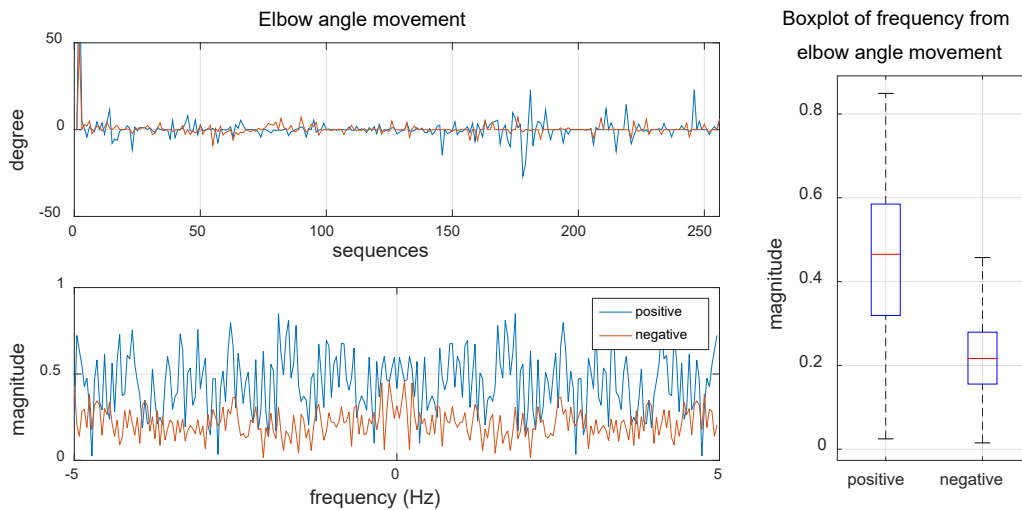


Figure 3-17 Elbow angle movement

All analysed data, from 28 clip videos gave similar results. Clearly, participants tend to move their hands frequently with high-quality movement when they are in a positive mood. The movements exhibit a high changing rate and high speed.

### 3.2.3.3 Bodily state of expression in mood scene

As noted above, the variety of expression data from videos can be represented as simple actions that fall into 4 main categories, namely, two posture forms, and two quality movements in actions. Each action sequence was categorized according to conditions obtained from the average behaviour as shown in the previous section with using Equations (3-1) – (3-4).

$$\text{raised form} = \begin{cases} \text{raised}, & d_y \geq 0.14 \text{ m.} \\ \text{normal}, & \text{otherwise} \end{cases} \quad (3-1)$$

$$\text{opened form} = \begin{cases} \text{opened} & l_2 \geq \text{shoulder width} \\ \text{neutral}, & 0.19 \text{ m.} \leq l_2 \leq \text{shoulder width} \\ \text{closed} & l_2 \leq 0.19 \text{ m.} \end{cases} \quad (3-2)$$

$$\text{gesture speed} = \begin{cases} \text{high}, & v \geq 0.39 \text{ m/s} \\ \text{low}, & \text{otherwise} \end{cases} \quad (3-3)$$

$$\text{direction changing} = \begin{cases} \text{high}, & \text{magnitude response} \geq 0.488 \\ \text{low}, & \text{otherwise} \end{cases} \quad (3-4)$$

Types of stationary posture can be distinguished by means of LMA-Shape conditions. So, the stationary state of bodily expression can be determined from the sub-states in Equation (1) and Equation (2). The "move" state can be specified using the conditions in Equation (3) and Equation (4). Hence, the continuous motion of the body can be decomposed into 10 distinct states. The list of states and their corresponding expressions is given in TABLE. 3-10.

Table 3-10 Lists of the states and their corresponding expressions

LMA-Features	States	Expression
LMA-effort (mobile)	A	High direction changing, Low speed
	B	High direction changing, High speed
	C	Low direction changing, Low speed
	D	Low direction changing, High speed
LMA-shape (immobile)	E	Closed form, Raised form
	F	Opened form, Normal form
	G	Neutral form, Raised form
	H	Closed form, Normal form
	I	Opened form, Raised form
	J	Neutral form, Normal form

The frequency of occurrence of each state, extracted from the narrative clip videos, were obtained by computing the number of times a state occurs within a narrative as shown in Figure 3-18, which shows observed sequences. For example, in a series of 10 positive mood expressions, State B occurred 4 times. So State B has a 40% chance of being triggered from a positive mood.

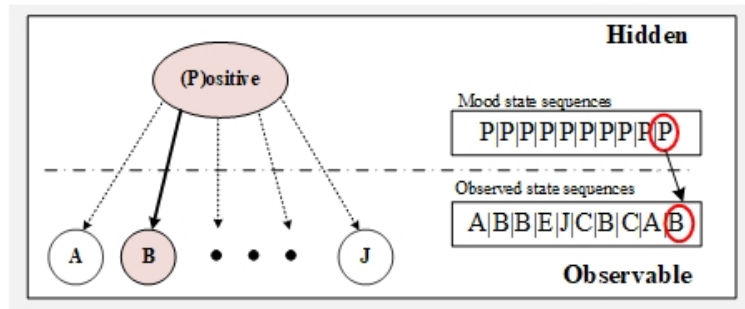


Figure 3-18 States Extracted from narratives

We analysed bodily-expression data afresh with state categorization. The rate of expression state occurrence, conditioned by mood, was obtained as shown in Figure 3-19. Clearly, participants tend to move their hands frequently with high-quality movement when they are in a positive mood (State B). The movements exhibit a high changing rate and high speed. By contrast, when participants were in the negative mood, they barely moved their hands (State C) and always held a closed posture (State E) more than other stationary states (F to J) while speaking.

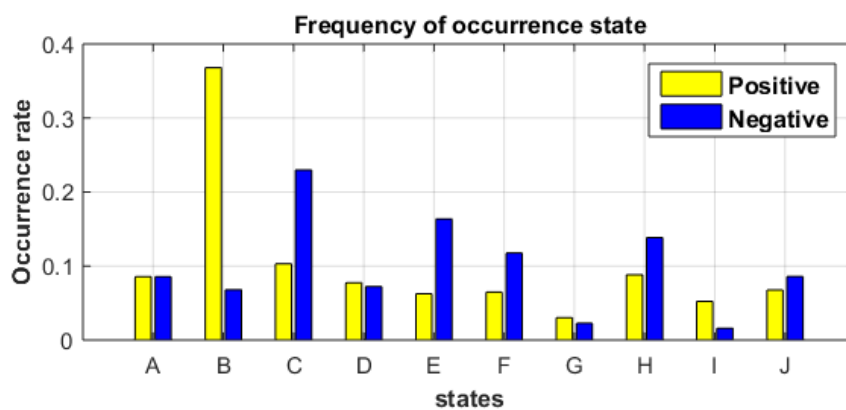


Figure 3-19 Frequency of occurrence state

### 3.3 Mood perception model using HMM

HMM are well known in pattern recognition system which becomes a widely used for interpreting, eliciting a hidden parameter of human activity, such as sign language recognition [28], activity recognition [29], even for recognizing the signal pattern of voice to estimate which words were spoken [30].

This work concept of mood perception model, assumes that changes in human moods can be represented as a discrete Markov chain but cannot be observed directly. Thus, we introduce a hidden Markov model. Mood strongly affects human behaviour, and feelings are normally conveyed by observable expressions. For example, one who feels angry in an interaction might try to intimidate or dominate the other person, and his or her posture would exhibit approach tendencies. From studied result in Section 3.2, It is reasonable to assume that narratives about positive and negative memories can be included in the same model. Therefore, mood perception using bodily movements can be linked to the hidden Markov model as shown in Figure3-20.

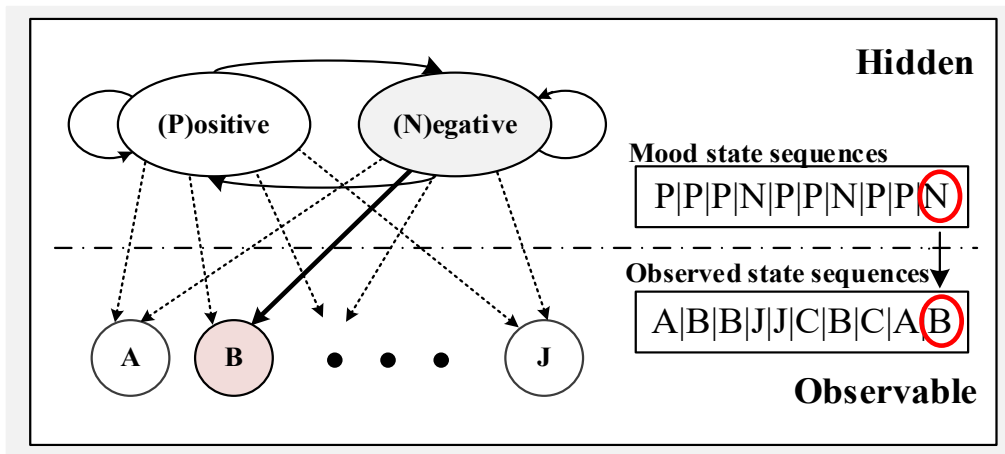


Figure 3-20 Bodily mood model

In this thesis, Mood perception model represents the new idea to create an effective mood perception system by using the bodily expression combine with facial expression. The occasioned of body states expression is used as emission probability from the hidden state (Mood) affect to

trig body expression. And, facial expression, the most significant indicator of emotional state, used to modify transition probabilities by means of the hidden mood state.

The convolution neural network (CNN) was adopted to create a model for recognizing emotion from a face image. The facial expression dataset (FER2013) was obtained from the website “Kaggle Facial Expression Recognition Challenge”, which consists of 28,709 training data, a validation set 3,589 examples, and a test set of 3,589 examples, with 48×48 pixel greyscale images of human face emotion, and 7 classes of face emotion (happy, surprise, neutral, sadness, fear, anger and disgust). The architecture of our CNN model for face emotion recognition is shown in Fig. 3-21.

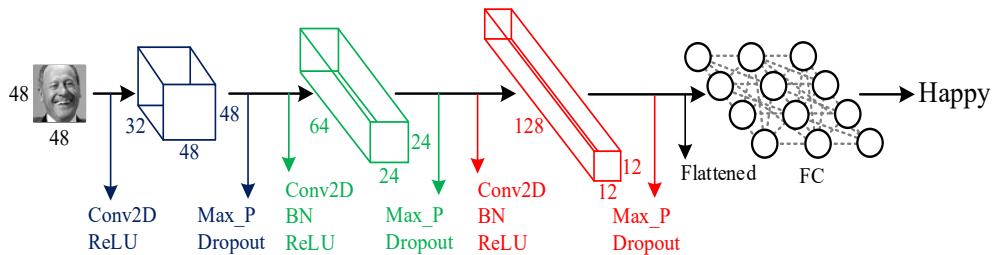


Figure 3-21 CNN model for face emotion recognition

The model consists of three convolution and three fully-connected layers. Each convolution layer is specified by 3×3 filters, batch normalization (BN), dropout, and max-pooling 2×2 filters. The output of the final convolution layer is flattened and connected to the fully-connected layer (FC). The details of the CNN architecture are given in TABLE 3-11.

Table 3-11 ARCHITECTURE OF CNN MODEL

Layers	Input	Output
Conv. 1	48×48×1	24×24×32
Conv. 2	24×24×32	12×12×64
Conv. 3	12×12×64	6×6×128
FC. 1(flattened)	4608	256
FC. 2	256	128
FC. 3	128	7 (emotions)



The model was trained with a batch size of 128 using the Adam optimizer to minimize the loss associated with “categorical cross-entropy.” The model's accuracy is 82.2725%. Note however that much of this data is superfluous for present purposes since positive or negative emotion alone provides enough information for the robot to choose the appropriate reaction when interacting with humans. So basic emotion (excluding “Neutral”) were classified into two groups: positive and negative emotional states. In Figure 3-21 shows output from the facial expression recognition program based on a CNN model. The most reliable score in Figure3-21 is always chosen as a single output class from the face emotion classes as shown in Figure 3-22.

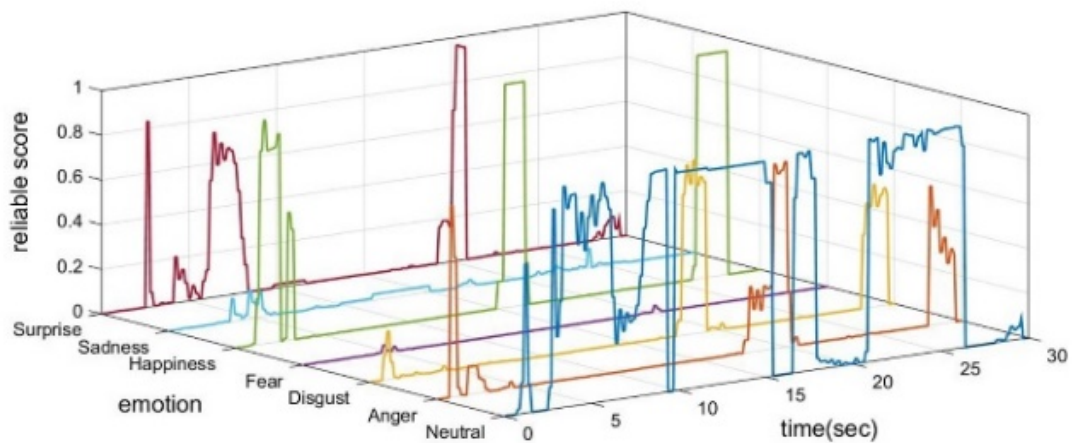


Figure 3-22 Emotion reliable scores



Figure 3-23 Face emotion classes

As noted in Section 3.2.1, “Mood” is a state of feelings, which contain complexity and multiple emotion. Thus, the transition probabilities in hidden layer are determined by computing the average prediction scores of both positive and negative groups of facial expression by using Equations (3-5), Equation (3-6), and Equation (3-7).

$$A = \begin{pmatrix} a_{p \rightarrow p} & 1 - a_{p \rightarrow p} \\ 1 - a_{n \rightarrow n} & a_{n \rightarrow n} \end{pmatrix} \quad (3-5)$$

$$a_{p \rightarrow p} = \frac{\sum_{x \in PE} x}{n_{PE}} \quad (3-6)$$

$$a_{n \rightarrow n} = \frac{\sum_{x \in NE} x}{n_{NE}} \quad (3-7)$$

The state transition matrix (A) shows the probability of a state change from time t to t+1, meaning the state remains the same or shifts to a different one. The probability of the next occurring state will happen to repeat, both positive ( $a_{p \rightarrow p}$ ) and negative ( $a_{n \rightarrow n}$ ), are computed from the average facial reliable score  $x$  in each emotion expression group. The study reported in [31] divides positivity and negativity into six basic emotion across cultures, “Anger, Fear, Sadness and Disgust” are in the negative group (NE); “Happiness” is the most significant of emotion in the positive group. “Surprise” is a more ambivalent emotion than others, but it has a stronger rating in positivity than negativity. In case of “Neutral” facial expression, the state transition probabilities of switching or staying in the same state are each 0.5. Furthermore, the frequency of occurrence of each

expression state, in narration, was used as an emission probabilities ( $e$ ), which represents how likely each bodily expression state was triggered by a hidden state, positive or negative mood. Figure 3-23 provides an overview of the model

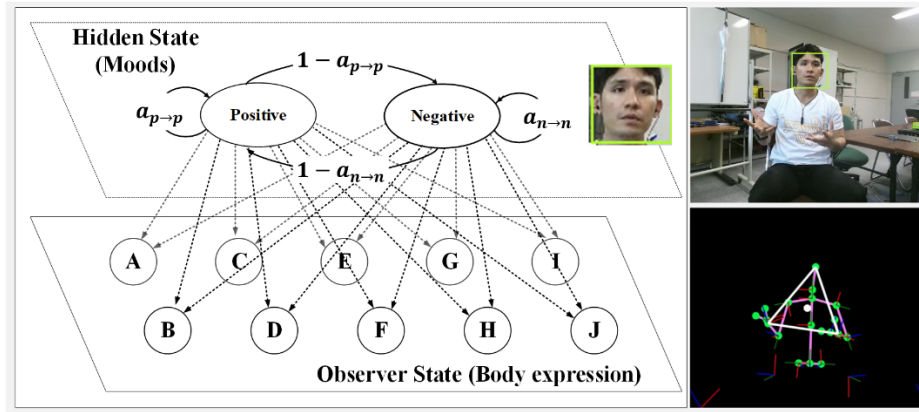


Figure 3-24 HMM of Moods perception model

Table 3-12 HMM parameter for Mood perception model

HMM	Model Parameters
Initial prob.	Initial probabilities: $\Pi = \{0.5, 0.5\}$
Hidden state sequences $X = (x_1, x_2, \dots, x_n)$ with $x_n \in S = \{s_1, s_2, \dots, s_K\}$	Mood state: $K=2$ ; $S = \{\text{Positive}(PE), \text{Negative}(NE)\}$
Observable sequences $Y = (y_1, y_2, \dots, y_n)$ with $y_n \in O = \{o_1, o_2, \dots, o_N\}$	Bodily expression state: $N=10$ ; $O = \{A, B, C, D, E, F, G, H, I, J\}$
Transition prob.	$A = \begin{pmatrix} a_{p \rightarrow p} & 1 - a_{p \rightarrow p} \\ 1 - a_{n \rightarrow n} & a_{n \rightarrow n} \end{pmatrix}$ $a_{ij}$ representing the probability of moving from hidden state $i$ to hidden state $j$
Emission prob.	$e = \begin{pmatrix} e_{PE,A} & e_{PE,B} & \dots & e_{NE,J} \\ e_{NE,A} & e_{NE,B} & \dots & e_{NE,J} \end{pmatrix}$ $e_{i,j}$ or $e(O_j   S_i)$ stores the probability of observing $o_j$ from state $s_i$

HMM model parameters for estimating human mood are represented in Table 3-11. The mood perception model consists of following parameters: ( $Y$ ) is the observation sequences from bodily actions; ( $A$ ) are transition probabilities inside hidden mood state; and ( $e$ ) are emission

probabilities from hidden state affect to observation state. Hidden states consist of two moods, positive and negative, that emit bodily expression state ( $O$ ) with emission probabilities ( $B$ ), obtained from the result from Figure 3-20, as shown in Equations (3-8),

$$e = \begin{pmatrix} 0.087 & 0.368 & 0.103 & 0.077 & 0.062 & 0.064 & 0.030 & 0.088 & 0.054 & 0.067 \\ 0.086 & 0.068 & 0.229 & 0.072 & 0.166 & 0.117 & 0.023 & 0.138 & 0.016 & 0.085 \end{pmatrix} \quad (3-8)$$

From Equation (3-8), if the participant has a positive mood, there is a 36.8% chance that he would exhibit state “B” action; if he has a negative feeling, there is a 13.8% chance that he would express state “H”.

### 3.4 Solving kinematic arm using NNs and backpropagation

This section presents the model for a 6 DOF Seamless robotic arm using the inverse kinematics solution based on artificial neural network (ANN). Levenberg-Marquardt method is used in training algorithm. The desired position and orientation of the end effector is defined as the input pattern of neural network.

#### 3.4.1 Kinematic of Seamless robot’s arm

We have considered and operated two robotic arms, but for demonstration in this paper, we present all equations for the left side of the robotic arm only. The first step in deriving a forward kinematics solution is to find the parameters of the Denavit-Hartenberg method. Figure3-24 shows the joints and links of the left arm, which has the base frame, shoulder, elbow, and wrist. The DH parameters of the robotic arm are listed in Table 3-12. According, DH method, the forward transformation matrix can be formed in the chain product of five homogenous matrices  ${}^{i-1}T_i$  as expressed by Equation (3-9).

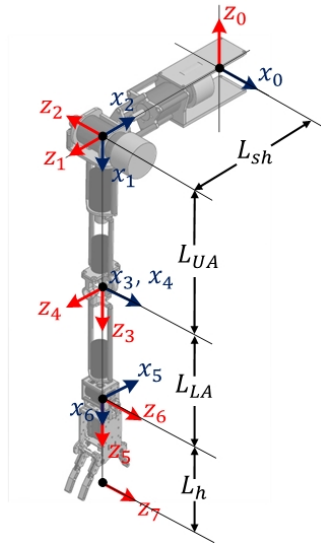


Figure 3-25 Frame assignment for Seamless arm

Table 3-13 DH parameter of Seamless Robot left arm

Joint (i)	DH Parameter			
	$\alpha_{i-1}$	$a_{i-1}$	$d_i$	$\theta_i$
1	$90^\circ$	0	$L_{sh}$	$\theta_1 - 90^\circ$
2	$-90^\circ$	0	0	$\theta_2 + 90^\circ$
3	$90^\circ$	0	$L_{UA}$	$\theta_3 + 90^\circ$
4	$-90^\circ$	0	$-90^\circ$	$\theta_4$
5	$90^\circ$	0	$L_{LA}$	$\theta_5 - 90^\circ$
6	$-90^\circ$	0	$90^\circ$	$\theta_6 - 90^\circ$
eef	0	$L_h$	0	0

$${}^0T_{eef} = {}^0T_1 {}^1T_2 {}^2T_3 {}^3T_4 {}^4T_5 {}^5T_6 {}^6T_{eef} \quad (3-9)$$

### 3.4.2 Neural Networks for solving kinematic problem

Solving the kinematics problem of robotics arm is one of the most important task in robot kinematics and control. The complexity of this problem is obtained from the robot's geometry and the nonlinear trigonometric equations. There are three classical techniques used for determining inverse kinematics problem that can be classified as algebraic, geometric and numerical algorithms. However, these methods are time consuming, incorrect initial perception,

and heavy computational calculation. Consequently, another method can solve the inverse kinematics problems that is backpropagation Neural Networks. It is possible to formulate the nonlinear mathematical model for solving robotic arms. Consequently, another method can solve the inverse kinematics problems that is Neural Networks. It is possible to formulate the nonlinear mathematical model. Many researchers have verified with this approach in several works [32], [33]. Thus, neural networks was adopt to solving robotic arm in this thesis, which the designed neural network as is illustrated in Figure 3-25.

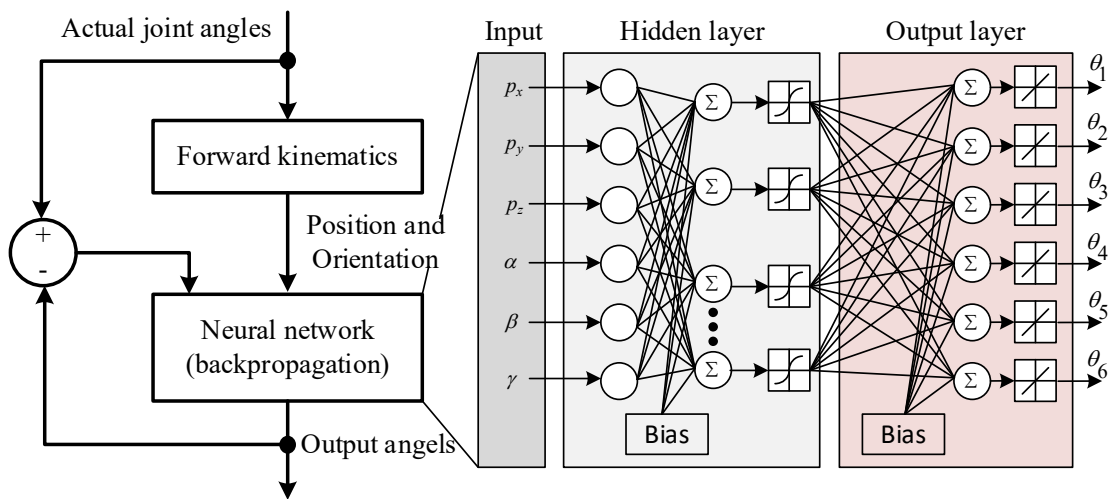


Figure 3-26 Kinematics solution based Neural Network

From diagram in Figure3-24, the set of operation angle (6 joints) was randomly generated and fed to forward kinematic. The result obtained from forward kinematic model and its actual joint will be used as dataset for training the desired model.

### 3.4.2 Robot gesture generation based human emotional expression

A robot with emotional expression capabilities could communicate more successfully in part of social interaction, as shown in [34]. Moreover, in the work [35], whose authors concluded that a robot in physical form is better suited for interaction with humans than a bodiless computer program. The ultimate goal of HRI is not only effective robotic perception, but also the ability of a robot to communicate information that humans can perceive. Thus, in this section, a method for

creating an emotional gesture of the Seamless robot will be described. In section 3.4.2.1, the triangle shape parameters, obtained from human emotional expression will be served as the basis for generating the robot gestures. Toward this end, in section 3.4.2.2 also propose a method for evolving robot emotional expression based on human evaluation. Experiments were conducted with the upper-body of a Seamless robot to verify and evaluate the effectiveness of the proposed system.

### 3.4.2.1 Emotional expression platform based on LMA-shape

LMA is not only used for interpret human behaviour in emotional expression, but also utilized for generate robotic expression system as reported in [16], [36], [37]. LMA-shape properties based on the studied result of human emotional expression is used to generate Seamless-robot emotional expression platform. In Figure3-26 shows the diagram of the gesture generation procedure.

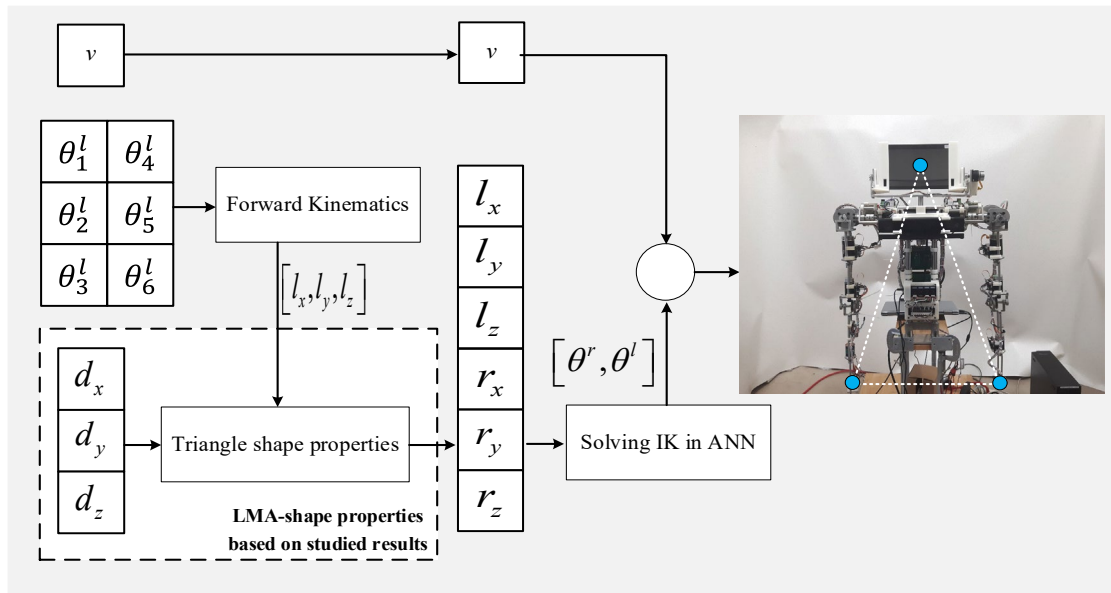


Figure 3-27 Gestures based on LMA model

First, the operation angles of the right-arm ( $\theta_1^l, \theta_2^l, \theta_3^l, \theta_4^l, \theta_5^l, \theta_6^l$ ), the centroid of triangle position ( $d_x, d_y, d_z$ ) and the rotational speed  $v$  are randomly chosen. Next, the operation angles

are entered into the forward kinematics equations in order to compute the right hand position  $(r_x, r_y, r_z)$ . Then, the centroid of the triangle, the right-hand position and the head position  $(h_x, h_y, h_z)$  are used to determine the left-hand position of the robot  $(l_x, l_y, l_z)$  which is computed using Equation (3-10). To control each arm, these two positions are entered into the ANNs kinematics model to obtain their operation angles.

$$\begin{aligned} d_x &= \frac{1}{3}(r_x + l_x + h_x) \\ d_y &= \frac{1}{3}(r_y + l_y + h_y) \\ d_z &= \frac{1}{3}(r_z + l_z + h_z) \end{aligned} \quad (3-10)$$

### 3.4.2.2 Interactive evolution computation

A Genetic Algorithm (GA) was adopted for developing and optimizing robotic emotional gestures. First, we arbitrarily chose to make the robot's left hand dominant, and initialized its movement with gesture generation, each gesture consisting of 3 segments of motion  $[m_1, m_2, m_3]$ . The *chromosome* has nine control parameters (*genes*), namely, six right-arm operation angles, three values defining the position of the triangle shape. The chromosome structure is illustrated in Figure3-27. The overall system diagram is presented in Figure3-28.

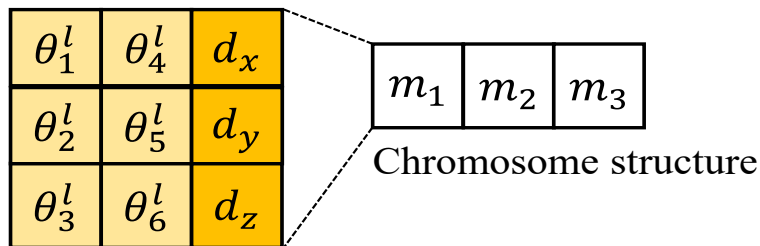


Figure 3-28 Chromosome structure



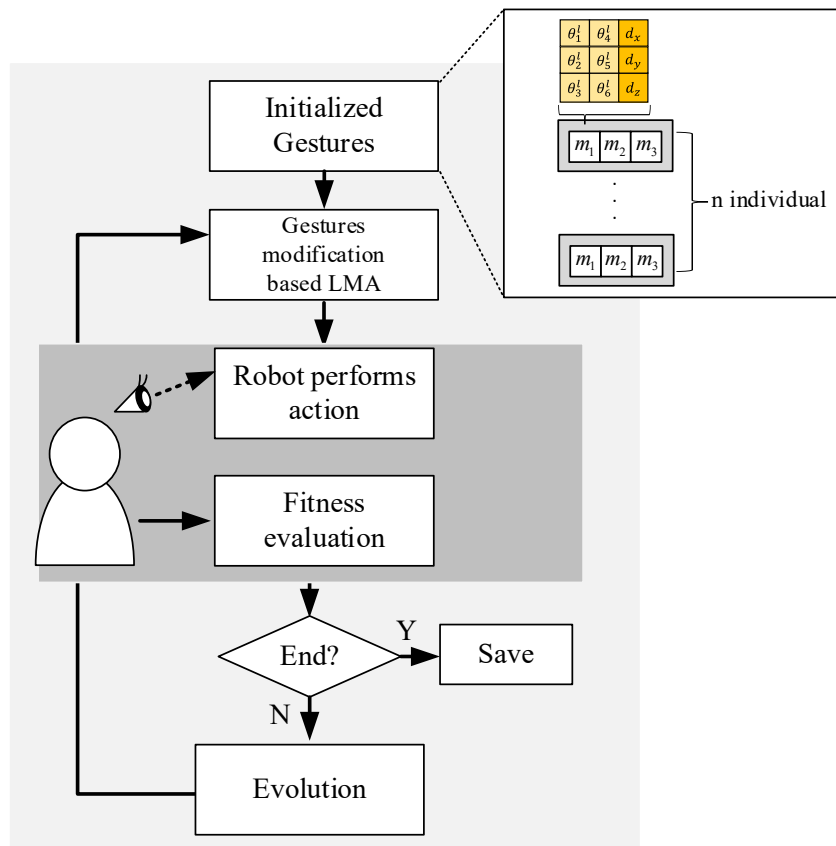


Figure 3-29 Overall gestures system diagram

After initialization, the gesture of the left arm is determined from the LMA-shape data. Then, each gesture is performed with both arms. However, evaluation of the fitness value of emotional expression is complex and difficult to analyse automatically. So, the typical fitness function in the GA model is produced by hand. Each gesture in the population is rated by a human. Then, the strongest individual is picked by means of tournament selection to produce the new offspring, and rated again as in the previous process. To quicken evolution, the weakest individual (unfit) in the population is replaced by the stronger one of the new offspring. However, this replacement process may wipe out population diversity. To prevent that from happening, the similarity between new candidate offspring and the weakest individual is checked before replacement. The procedure for this human-interactive evolution is shown in the following pseudo code.

Table 3-14 Procedure of human-interactive

<b>Evolutional for robot emotional expression</b>
1: Function Evolving(population):
2: select the strongest parents
3: produce new offspring as standard GA
4: robot express emotional gestures of offspring
5: user rates the performance of robot
6: if offspring fitness > weakest fitness in population
7:     replace weakest population with offspring
8: end if
9: while average fitness is lower than stop condition
10:    for all gestures do
11:       robot express emotional gestures;
12:       user rates the performance of robot;
13:    end for
14:    evolving(population);
15: end while
16: save gestures pattern;
17: end function

### 3.5 Summary

In this chapter, the configuration of the robot head and arm were discussed. And also the explanation of control system. The implementation of neural networks for solving the robotic arm is also explained. Followed with the study process for understanding human-emotional expression, with two moods expressional scheme in both good and bad feelings. The methodology in this chapter allowing this thesis to create a robot gesture generation to express emotional that human can perceive. Moreover, the combination of facial and bodily expression in determining human emotional states are also represented.

## Chapter 4

### Experimental results

This chapter reports two experimental results, which, observations of human emotional expressions have been taken as a model for creating the robot emotional interaction system. This system allows the robot to communicate information that humans can perceive, and also enhance a new recognition ability to understand human emotion. With this 'full duplex' capability, HRI will become truly effective.

The data thus obtained was analysed by means of the Laban Movement Analysis (LMA) method to elicit critical parameters. The first experiment was designed for generating corresponding robotic movement. This experiment was conducted with the upper-body of a Seamless robot to verify and evaluate the effectiveness of robot emotional expression system to human perceptions. The second experiment showed the result of mood perception model, followed with the comparison result between the purposed model and human perception.

#### 4.1 Experiment I: Seamless robot emotional expression

Two distinct mood (Positive, and Negative) are used to verify the effectiveness of the proposed emotional expression schemes (Figure3-26). Each emotional scheme has 10 sets of emotional expressions, which consists of two expression sets. The first one relates to joint arm movements which are randomly generated to express the desired emotion. The consistency between arm movement and target emotion is evaluated by a human observer, and the best fit movement is used to realize the best gestures. By contrast, eye and head movements are expressed in terms of emotion schemes with predefined gestures. The initial gestures is shown in Table 4-1, and the final generation of both positive and negative emotional expression in Table 4-2, 4-3. Fig 4-1 shows the results of the evolution of the fitness value, which confirmed the emotions of the robot based on human interaction in a given situation.

Table 4-1 Initial gestures


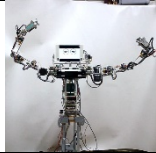
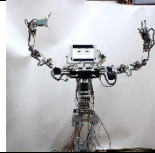
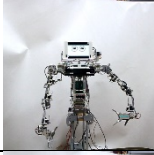
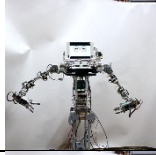
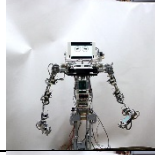
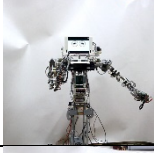
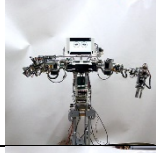
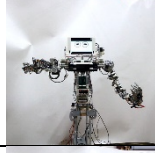
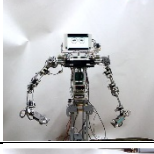
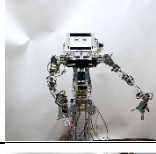
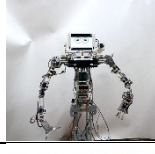
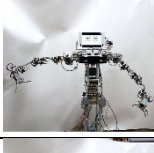
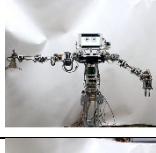
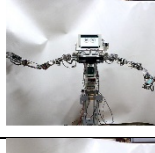
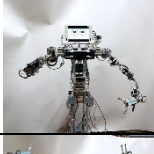
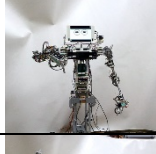
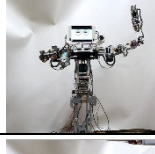
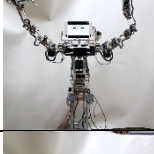
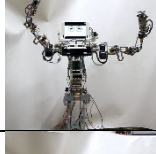
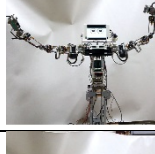
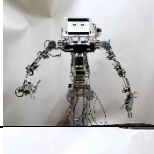
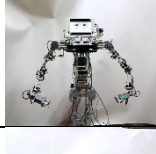
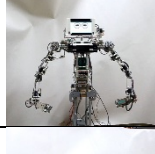
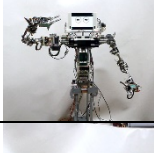
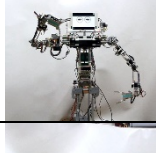
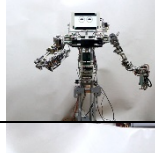
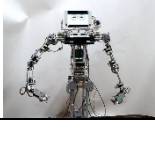
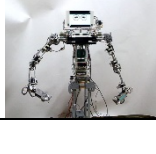
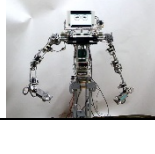
Population	$m_1$	$m_2$	$m_3$
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

Table 4-2 Final generation of positive expression





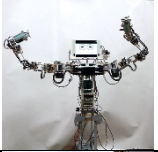
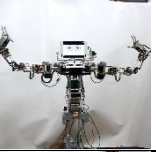
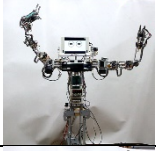
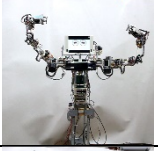
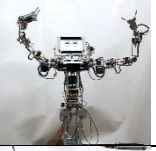
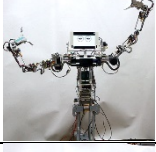

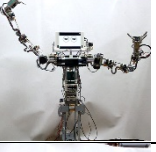
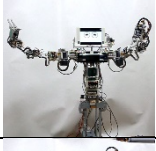
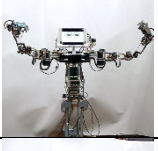
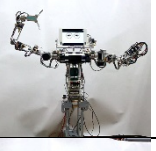
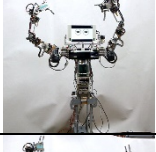
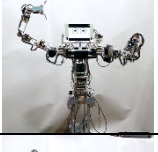

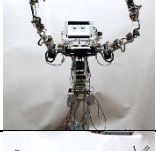
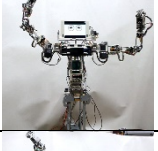
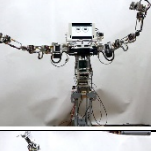
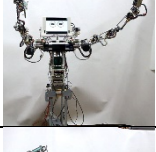
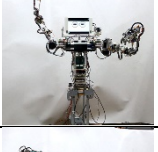
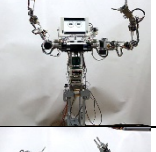
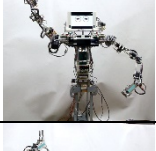
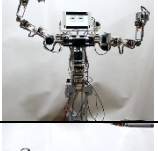
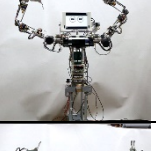
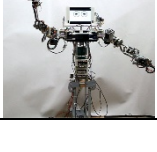

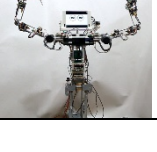





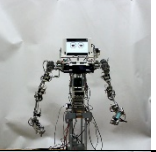

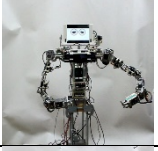

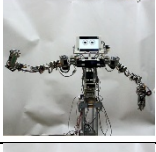


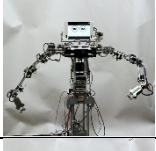
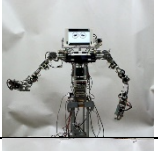
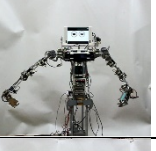
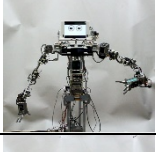
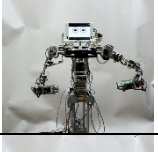

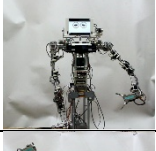

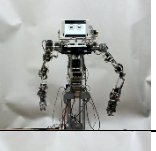
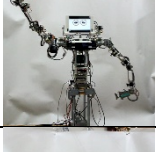

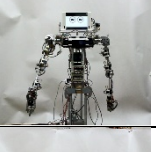
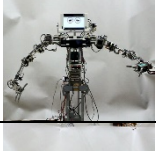
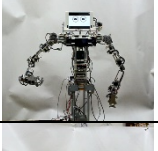
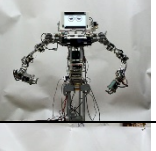

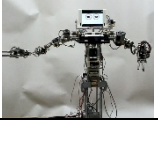
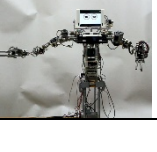
Population	$m_1$	$m_2$	$m_3$
1 Fitness = 80			
2 Fitness = 90			
3 Fitness = 90			
4 Fitness = 90			
5 Fitness = 75			
6 Fitness = 75			
7 Fitness = 80			
8 Fitness = 75			
9 Fitness = 78			
10 Fitness = 80			

Table 4-3 Final generation of negative expression

Population	$m_1$	$m_2$	$m_3$
1 Fitness = 80			
2 Fitness = 90			
3 Fitness = 80			
4 Fitness = 80			
5 Fitness = 80			
6 Fitness = 90			
7 Fitness = 75			
8 Fitness = 70			
9 Fitness = 75			
10 Fitness = 80			

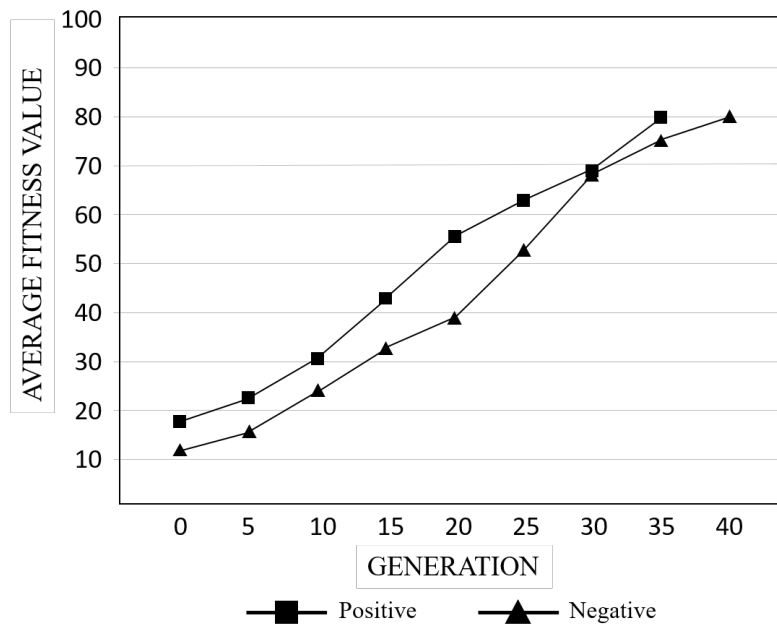


Figure 4-1 Evolution of fitness value

The evolution of fitness value in Figure 4-1, have shown that Seamless robot can learn and adapt its emotional gestures based on human evaluation. The highest centroid position,  $d_y=0.546m$ , corresponds to a positive gesture, while the lowest one is in negative scheme, with  $d_y=0.159m$ . These experimental result also provide strong evidence that humans tend to perceive the other emotions by implicitly interpreting the distance between the centroid position and each plane ( $d_x, d_y, d_z$ ). Our main objective is to develop a robot with high potential in interactive skills.

## 4.2 Experiment II: Mood perception system

The experimental data has been obtained by estimating human moods from two narrative scenes, which were shared by participants. The sequences of expression data were input to the model every 100ms. First, face image data was sent to the CNN model for updating transition probabilities between positive and negative states. For testing the accuracy of the system, the narration data were fed to the model for estimating the mood of the narrator. Each expression state was labelled according to its mood scene. The Viterbi algorithm was used to analyse the

most likely sequence, and the mood result of each expression state was compared to its actual label. The number of correct predictions was used to describe the accuracy of the model as shown in Figure 4-2.

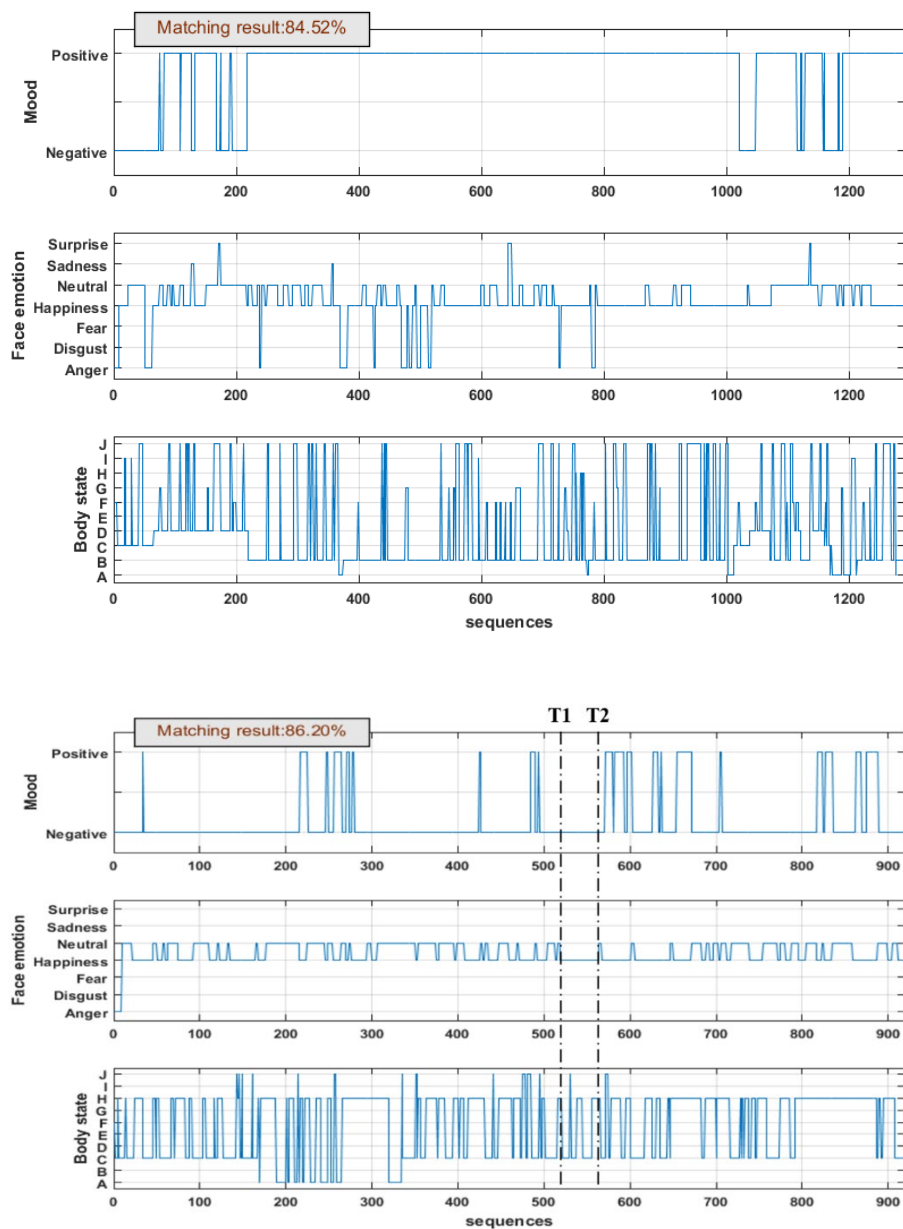


Figure 4-2 Mood analysed result  
 Top: Positive experience; Bottom: Negative experience

Figure 4-2 illustrated mood estimation result from both positive (top figure) and negative scene (bottom figure), each figure illustrated mood analysed result (top graph) from both



observable information, face (middle graph) and body expression (bottom graph). Figure 4-2 (top) shows the result obtained from positive narration story, namely, that more than 84.52% of sequences are estimated as positive moods. The results suggest that while sharing this story, this person was in a positive mood. The results in Figure 4-2 (bottom) shows that 86.20% of sequences were occasioned by negative hidden mood states. Moreover, the facial expression gave an interesting result, i.e., mood results show that sometimes happy faces do not always express happiness, and the participants tend to express “Neutral” face in general conversation more than other intense emotions either in "Positive" or “Negative” group of expression. In Figure 4-3 illustrated the result of mood estimation accuracy by two different method, facial combining with bodily expression, and facial expression alone. The comparison of accuracy results of positive scenes is slightly different. Nevertheless, the accuracy result of facial combine with bodily expression for estimating negative mood are outstanding.

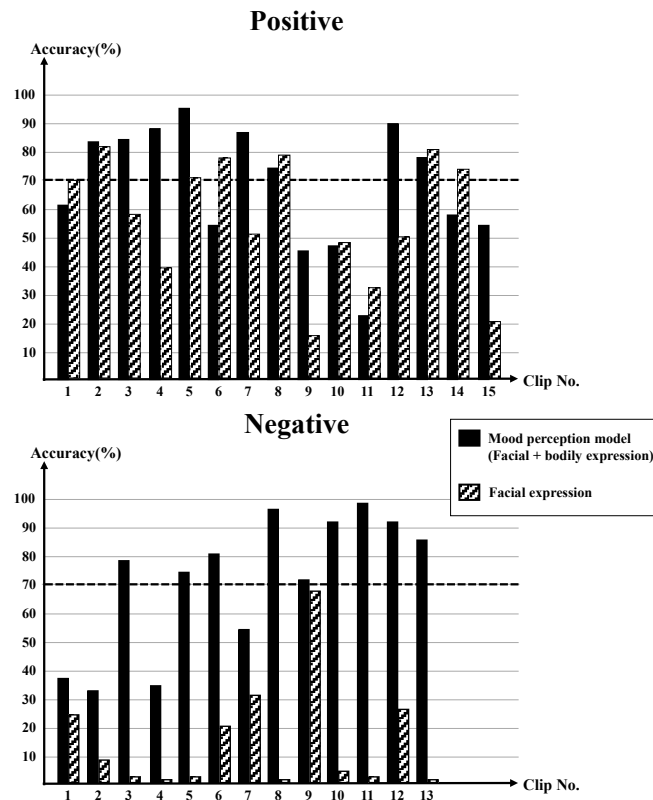


Figure 4-3 Comparison of accuracy result by using two different input  
 Top: Positive scenes; Bottom: Negative scenes

### 4.2.1 Comparative accuracy of Mood estimation

To confirm that the mood perception model could work as well as a human observer, we asked the other group of participants, 42 male, and 133 female (18-30 years old), to estimate mood by watching the same video clips (without sound) as we used for testing the model accuracy. They could see only the faces and bodily behaviours of the narrators. The number of correct answers was used to calculate the accuracy of human perception. The comparison of accuracy results between our perception model and human perception is shown in Figure 4-4.

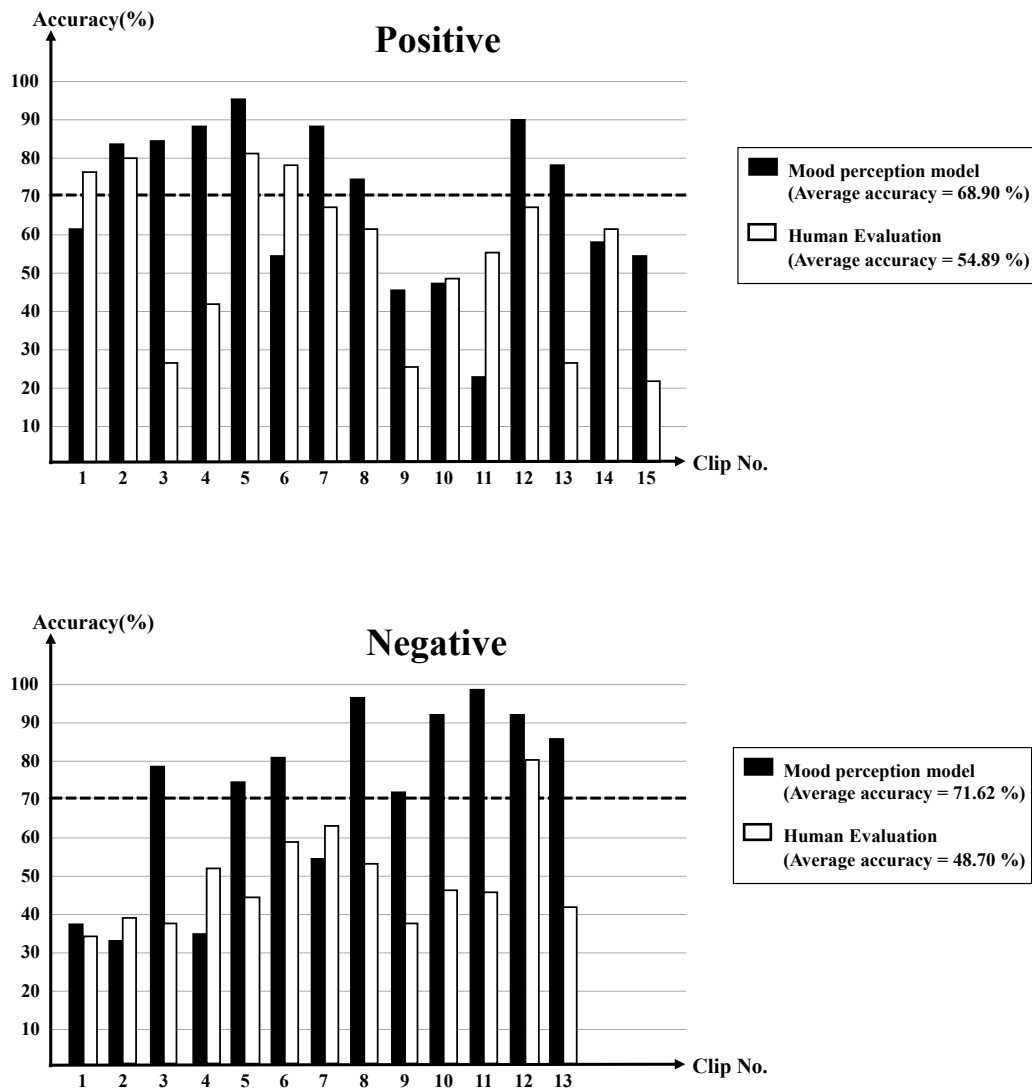


Figure 4-4 Comparison of accuracy result between mood model and human evaluation  
Top: Positive scenes; Bottom: Negative scenes

Moreover, the accuracy results from the purposed model and human evaluation were analysed by a statistical method called Paired t-test. The paired t-test is a statistical procedure used to determine the difference of scores between two methods of measurement (Mood perception model and Human evaluation), where each method is applied to the same subjects (Narrative clips). Like many statistical procedures, the paired t-test has two competing hypothesis testing, the null hypothesis and alternative hypothesis. To confirm that there is no difference between two method being studied, the comparative accuracy was test under null hypothesis.

The comparative accuracy data from figure 4-3 were computed by statistic equation in Equation (4-1) – Equation (4-3).

$$\bar{d} = \frac{\sum_{i=1}^n d_i}{n} \quad (4-1)$$

$$S_D = \sqrt{\frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n - 1}} \quad (4-2)$$

$$t = \frac{\bar{d} - \mu_d}{\frac{S_D}{\sqrt{n}}} \quad (4-3)$$

Where:

- $n$ : Sample size
- $d_i$ : The difference between two method on each pair
- $\bar{d}$ : The mean of difference of  $d_i$
- $S_D$ : The standard deviation of differences
- $t$ : t-statistic

For conclusion about the difference in performance of mood perception between human evaluation and the purposed model, the accuracy information from Figure 4-3 : Positive results ( $n=15$  clip videos) and Negative result ( $n=13$  clip videos) were test under null hypothesis condition. The null hypothesis assumes that the true mean difference is equal to zero ( $\mu_d= 0$ ). The

paired t-test calculation from positive clip videos and negative clip videos can be represented in Table 4-4, and Table 4-5, respectively.

Table 4-4 Paired t-test calculation from Positive clip videos

Clip	Purposed model (x)	Human evaluation (y)	$d_i = x_i - y_i$	$d_i - \bar{d}$	$(d_i - \bar{d})^2$
1	61.080	77.143	-16.063	-30.068	904.062
2	83.400	80.000	3.400	-10.605	112.461
3	84.520	26.286	58.234	44.230	1956.251
4	88.390	47.714	46.676	32.671	1067.391
5	96.310	80.571	15.739	1.734	3.006
6	55.710	78.286	-22.576	-36.580	1338.131
7	88.040	67.429	20.611	6.607	43.648
8	73.310	62.857	10.453	-3.552	12.616
9	47.430	26.286	21.144	7.140	50.973
10	48.560	49.143	-0.583	-14.588	212.799
11	23.110	55.429	-32.319	-46.323	2145.851
12	90.670	68.000	22.670	8.665	75.086
13	79.910	27.429	52.481	38.477	1480.454
14	58.960	60.571	-1.611	-15.616	243.865
15	54.100	22.286	31.814	17.810	317.179
			$\bar{d} = 14.0047$		<b>SUM=9963.773</b>

The standard deviation of accuracy differences of positive clips video was computed by Equation (4-2) using data Table 4-4 ( $S_D = 26.677$ ), then  $t$ -value ( $t = 2.03$ ) was obtained from Equation (4-3). Then, calculate  $p$ -value from our calculated  $t$ -value in T-distribution table (see in Appendix B). A  $p$ -value is used to determine whether the null hypothesis should be rejected or retained: The

null hypothesis is rejected if the p-value is less than (or equal to) the alpha level ( $\alpha$ ), the alpha level is usually set at or below 5% ( $\alpha = 0.05$ ), which conventionally refer to statistically significant.

The null hypothesis is retained because p-value from the paired t-test result of the positive clip video (t-value = 2.03, p-value = 0.061), is more than alpha level (0.05). It means there is no different in perception accuracy of positive data between two method of measurement.

Table 4-5 Paired t-test calculation from Negative clip videos

Clip	Purposed model (x)	Human evaluation (y)	$d_i = x_i - y_i$	$d_i - \bar{d}$	$(d_i - \bar{d})^2$
1	38.080	34.857	3.223	-30.068	904.062
2	31.310	38.857	-7.547	-10.605	112.461
3	78.600	36.000	42.600	44.230	1956.251
4	34.090	52.000	-17.910	32.671	1067.391
5	74.970	44.571	30.399	1.734	3.006
6	80.530	59.429	21.101	-36.580	1338.131
7	53.720	63.429	-9.709	6.607	43.648
8	96.310	53.714	42.416	-3.552	12.616
9	70.780	37.143	33.637	7.140	50.973
10	92.950	46.286	46.664	-14.588	212.799
11	99.920	45.714	54.206	-46.323	2145.851
12	93.800	80.000	13.800	8.665	75.086
13	86.200	41.143	45.057	38.477	1480.454
			<b><math>\bar{d} = 22.918</math></b>		<b>SUM=7105.230</b>

For the negative clip, there was a significant average difference between mood the perception model and human evaluation (t-value = 3.40, p-value = 0.005). In any case, including bodily expression in the analysis, our perception model gave a better result with accuracy over 70% correct perception, for up to 17 video clips. Only 5 correct results were given human observers. In general,

these results show that a mood perception model combining both facial and bodily expression is useful for recognizing human mood states.

#### 4.2.2 Human-robot emotional interactive task

Additionally, this thesis also demonstrates the use of mood perception in interaction between human and robot. The work flow diagram of the interactive system is shown in Fig 4-5. This interactive system allows the robot to communicate with user with full human perception skills.

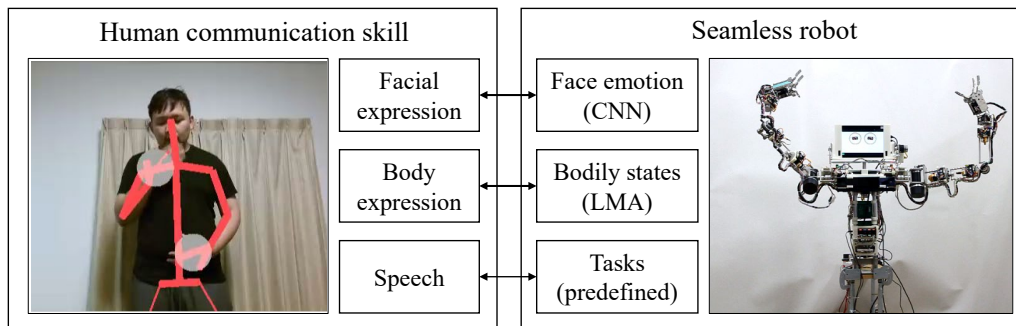


Figure 4-5 Interactive system

In this thesis, the voice command was predefined, with the simple phrase, in order to control the robot to do the desired task. The detail of the command and its task are shown in Appendix A.

To demonstrate how this research can be used as for the HRI system, the mood perception model is applied to create a simple interactive game between human and robot, called "Guessing game". This game shows that human mood can be used as additional input, for enhancing the social robot skill. Figure4-6 illustrates the Guessing game. The game requires the robot to drop an object into a box chosen by the human participant, taking account of the participant's reaction.

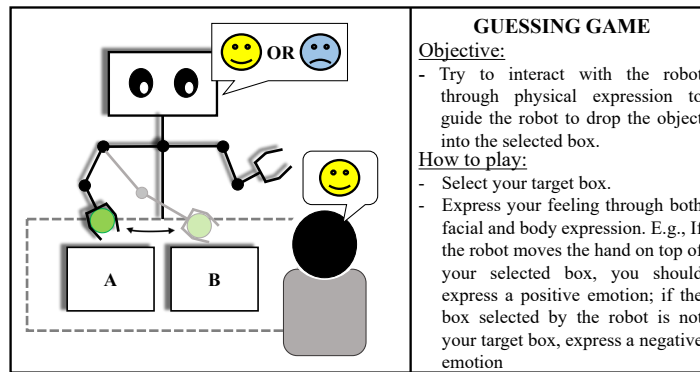


Figure 4-6 Infographic of guessing game

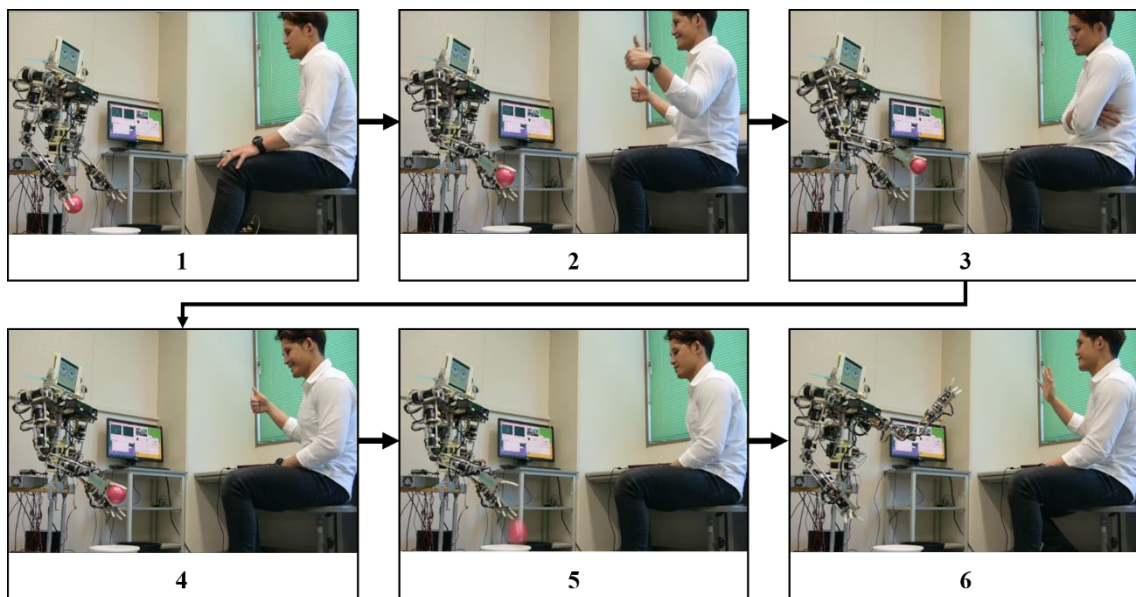


Figure 4-7 Seamless robot and human performed Guessing game

In Figure 4-7, the game starts when the robot grips the ball in its hand (Pic. 1); then the robot moves its hand over two boxes. During the robot jump its hands switch back and forth over two boxes, and the human participant's mood is then analysed (Pic.2 to Pic.4). If the perception result signals positive mood, it is assumed that the participant wants the object in that box. The score of this game is calculated from the mood perception result while playing “Guessing game”. In the case of positive mood result, the score will be increased, and, decreased otherwise. When the score associated with a given box reaches a threshold (now arbitrarily set at 80 points), the robot will drop the object into that box (Pic. 5), and the game is over. The phrase "Bye-Bye" says good bye to the robot, in which case

the robot waves its hand (Pic. 6) and say goodbye back to the participant. Figure4-7 shows the mood perception program in operation while the participant is interacting with the robot

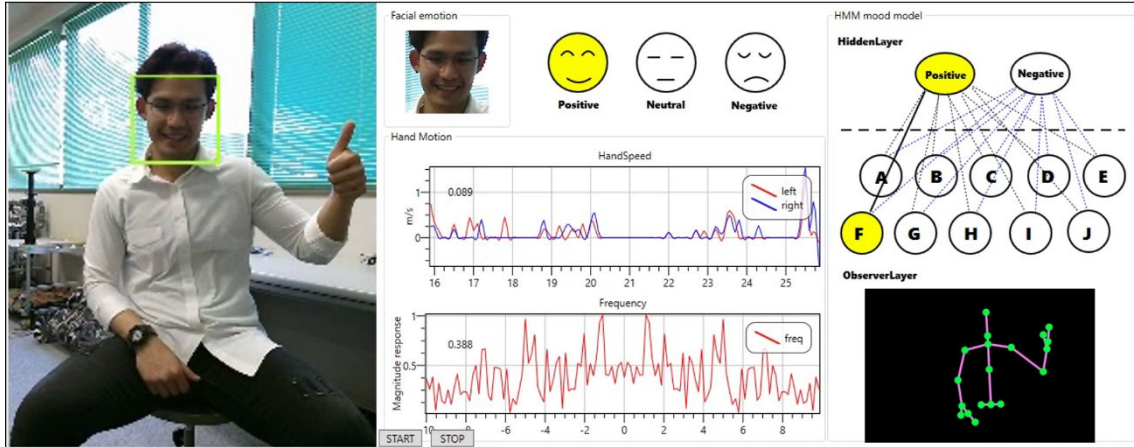


Figure 4-8 Mood perception program display

### 4.3 Summary

In this chapter, the proposed system is executed by the Seamless robot working along with human. The experimental results confirmed the effectiveness of the proposed system, which were represented by two experimental. In experimental I, the data of human expression is used to generate robot gestures, in order to allow the robot to communicate the emotion that human can perceive. LMA theory is adopt to explain and eliciting bodily parameters to clarify emotional expression more clearly. This experiment shows that LMA-shape and evolution computation improving the performance of robot emotional expression to be more fit to human perception. The robot perception system is also upgraded. The bodily expression data is used as an additional channel to create the essential innovation, HMM is adopt to combine two nonverbal input, facial and bodily expression, for estimating human mood states as illustrated in Experiment II. The estimating of the proposed model showed that mood perception model could work as the same sense as human. Moreover, the task “Guessing game” showed the benefit of using a mood perception model for intensifying perception skill of the social robot.



## **Chapter 5**

### **Conclusions**

#### **5.1 Conclusions of the research**

This thesis has presented an implementation of new skill for social robot, which is emotion interactive skill. The work attempts to use human physical behaviour during communication, facial and bodily expression, a critical communication skill to elicit unobservable “feelings”. This work purposed a method to collect the high reliable emotional expression data. The data were collected from the participants while they were sharing their own experience. The analysed data are used to model emotional interactive system, both expression and perception for robotic system. The result from a model of emotional expression evolution for the upper body of a robot has shown that Seamless robot can learn and adapt its emotional gestures based on human evaluation. This result provide a strong evidence that humans can perceive the emotions by robot physical movement.

The essential innovation in this approach is the combination of facial and bodily expression in determining emotional states. The advantage of this combination is clearly shown by the fact that, as the results have shown, facial expression alone may be misleading, since it may not be a true expression of inner feeling. However, combining facial with bodily expression, inner feelings can be estimated more reliably. Our main objective is to develop a robot with high potential in interactive skills. All objectives are achieved by the proposed system and they are verified by the experimental results in Chapter 4

The emotional interactive system is modelled into the social robot (Seamless robot) based on the studied of human emotional expression. The proposed system can be summarized by dividing to five chapters as described below:

Chapter 1 explains with the background history of the human and robot interaction, and the problem statement. The research target aims to develop a new communication ability for social

robot, with the more suitable interactive skill to work along with human. In order to achieve this goal, human behaviour in emotional expression was studied, then applied fundamental data to create an interactive model for a robotic system to imitate a human communication skill. And, at the end of chapter 1, the objectives of the thesis were specified.

Chapter 2 describes associated theories for this research. The "Laban movement analysis" (LMA) is adopted for obtaining bodily movement parameters in order to understand the relation between gesture and emotion of human. Then, the analysed data from LMA is then interpreted by "Hidden Markov Model" (HMM) combine with facial-emotional result from "Convolution Neural Networks" (CNNs). And, the robot emotional expression model is created by "Genetic Algorithm (GA)" based on the studied results from LMA. Then, Neural Networks was used for controlling and solving the kinematic problem of robotic arms. All the mentioned theories are explained briefly in this chapter.

Chapter 3 discusses the research methodology, consist with three main process, which are, the robot hardware and controlling design, Study of emotional expression, and the concept idea of mood perception model. In order to develop a kind of social robot, the robot was designed with functionality comparable to the human upper body consisting of trunk, head and two bilateral arms. Upper-body expression of the robot has two elements, namely, head and arm. in this chapter purposed a method for collecting the reliable data of human emotional body expression, and how to reducing complexity in human motion into a simple body state. The body state combines with facial emotion will be analysed by HMM to elicit unobservable mood state of human. Moreover, fundamental information from the studied result is also used to create robot emotional expression platform. Accordingly, all of the method described in this chapter allow the robot to perform full-duplex of emotional communication skill.

Chapter 4 shows experimental results from of the purposed research. The result is represented by two experiments. The first experiment presents an emotional expression platform for social robot, which was conducted with the upper-body of a Seamless robot to verify and

evaluate the effectiveness of robot emotional expression system. The result has shown that Seamless robot can learn and adapt its emotional gestures based on human. On top of that, the result of robot gestures can confirm that a human tends to receive the others' feeling through LMA-Shape parameters.

In second experiment, the studied result of bodily movement based on human emotion was involved to model a system to estimate human feeling. The model was developed by combining the result of bodily motion with facial expression, which is well known as the most distinct expression that conveys emotion. The output of the perception model is the simple human mood state since positive or negative emotion provides enough information for the robot to choose the appropriate reaction when interacting with humans. The proposed model was verified by comparing the result with human perception. The comparison results between the mood perception model and human are performed equally well, as regards accuracy. However, the mood perception results of the proposed thesis, by using a facial expression with bodily state expression as additional input, gave a better estimate result than human perception since human might excessively focus on reading emotion through facial expression. Moreover, the example of using mood perception model with social robot was also introduced by a simple task in this chapter. With this new capability in emotional interaction, it will allow the robot to be more human-like when working along with human.

Lastly, in this chapter, all the tasks of the proposed thesis were summarized. The objective of the thesis is to study and take advantage of human bodily motion as additional input for developing an emotional interactive system for social-robot. The emotional expression platform was conducted with Seamless robot, which was designed with functionality comparable to the human upper-body. With human upper-body mechanism, it enables the Seamless robot to take the studied data to be used as a platform to imitate human expression. On top of that, the mood perception model has shown that this model can be installed to the robotic system to enrich an interactive social system. Or, it could be applied in other human-computer interaction tasks. If a robot had

the ability to 'understand' human feelings, it would be able to communicate more effectively. The robot could use the subtle language characteristic of human speech to express observations. For example, if a person is in a bad mood from having failed in some activity, the companion robot might say **“You may need to spend a little more time practicing”** rather than **“You must improve your skill. You need to spend more time working on it.”** With the capability of mood perception, the companion robot would be able to conduct a more natural conversation.

## **5.2 Recommendations for future research**

The result of this thesis has been examined by a small group of participants. Anyway, this proposed thesis can nonetheless be considered as the first step in the development of emotional assessment based on human expression. To improve the performance of the system, the recommendations for future research are suggested as follow:

1. Nonverbal communication (facial, bodily) should be a further study to find a better ground truth among several humans.

2. Even though verbal languages is complicated, but either the meaning of the spoken words or tone of voices should be considered as one of the inputs for analysing human emotion too. With full human communication input (Verbal and Nonverbal language), the future version of emotional perception model is might be a high-performance model.

## References

- [1] A. Mehrabian, *Silent messages*, 1st ed. Belmont, Calif. : Wadsworth Pub. Co., 1971.
- [2] J. Li and M. Ousslah, “Automatic Face Emotion Recognition System,” *IEEE 9th Int. Conf. Cybern. Intell. Syst.*, pp. 1–6, 2010.
- [3] H. L. Ma Xiaoxi, Lin Weisi, Huang Dongyan, Dong Minghui, “Facial Emotion Recognition,” *Interfaces (Providence)*, 2017.
- [4] G. Castellano, S. D. Villalba, and A. Camurri, “Recognising Human Emotions from Body Movement and Gesture Dynamics,” pp. 71–82, 2007.
- [5] H. G. Wallbott, “Bodily Expression of Emotion,” *Eur. J. Soc. Psychol.*, vol. 28, no. 6, pp. 879–896, 1998.
- [6] A. Aviezer, Hillel; Trope, Yaacow; Todorov, “Body cues, not facial expressions, discriminate between intense positive and negative emotions,” *Science (80-. )*, vol. 338, no. November 2012, pp. 1225–1229, 2012.
- [7] N. Kubota and S. Wakisaka, “An Emotional Model Based on Location-Dependent Memory for Partner Robots,” *J. Robot. Mechatronics*, vol. 21, no. 3, pp. 317–323, 2016.
- [8] W. Jitviriyaya, M. Koike, and E. Hayashi, “Emotional model for robotic system using a self-organizing map combined with Markovian model,” *J. Robot. Mechatronics*, vol. 27, no. 5, pp. 563–570, 2015.
- [9] F. Jimenez, T. Yoshikawa, T. Furuhashi, and M. Kanoh, “Effects of a novel sympathy-expression method on collaborative learning among junior high school students and robots,” *J. Robot. Mechatronics*, vol. 30, no. 2, pp. 282–291, 2018.
- [10] D. B. Skillicorn, R. Billingsley, and M. Williams, “The Design Space of Social Robots,” 1979.
- [11] J. Denavit and R. S. Hartenberg, *A Kinematic Notation for Lower-pair Mechanisms Based on Matrices*. ASME, 1955.

- [12] W. Rawat and Z. Wang, “Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review,” *Neural Comput.*, vol. 29, pp. 1–98, 2017.
- [13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, *ImageNet Classification with Deep Convolutional Neural Networks*, vol. 25. 2012.
- [14] D. Chi, M. Costa, L. Zhao, and N. Badler, “The EMOTE model for effort and shape,” *Proc. 27th Annu. Conf. Comput. Graph. Interact. Tech. - SIGGRAPH '00*, no. July, pp. 173–182, 2000.
- [15] J. Morita, Y. Nagai, and T. Moritsu, “Relations between Body Motion and Emotion : Analysis based on Laban Movement Analysis,” pp. 1026–1031, 2001.
- [16] M. Masuda, S. Kato, and H. Itoh, “A Laban-based approach to emotional motion rendering for human-robot interaction,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6243 LNCS, pp. 372–380, 2010.
- [17] W. H. Kim, J. W. Park, W. H. Lee, M. J. Chung, and H. S. Lee, “LMA based emotional motion representation using RGB-D camera,” *ACM/IEEE Int. Conf. Human-Robot Interact.*, pp. 163–164, 2013.
- [18] M. M. Gross, E. A. Crane, and B. L. Fredrickson, “Effort-Shape and kinematic assessment of bodily expression of emotion during gait,” *Hum. Mov. Sci.*, vol. 31, no. 1, pp. 202–221, 2012.
- [19] E. A. Crane and M. M. Gross, “Effort-Shape Characteristics of Emotion-Related Body Movement,” pp. 91–105, 2013.
- [20] M. Gerber, T. Kaufmann, and B. Pfister, “Extended Viterbi algorithm for optimized word HMMS,” in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 4932–4935.
- [21] H. Zen, K. Tokuda, and T. Kitamura, “A VITERBI ALGORITHM FOR A TRAJECTORY MODEL DERIVED FROM HMM WITH EXPLICIT RELATIONSHIP BETWEEN STATIC AND DYNAMIC FEATURES Heiga Zen , Keiichi Tokuda ,

- Tadashi Kitamura Department of Computer Science and Engineering , Nagoya Institute of Technology,” no. 2, pp. 2–5.
- [22] M. Piórkowska and M. Wrobel, “Basic Emotions,” 2017.
- [23] I. Behoora and C. S. Tucker, “Machine learning classification of design team members’ body language patterns for real time emotional state detection,” *Des. Stud.*, vol. 39, pp. 100–127, 2015.
- [24] J. Inthiam, E. Hayashi, A. Mowshiwitz, and W. Jitviriyaya, “Development of an Emotional Expression Platform based on LMA-Shape and Interactive Evolution Computation,” pp. 11–16, 2018.
- [25] M. M. Gross, E. A. Crane, and B. L. Fredrickson, “Methodology for Assessing Bodily Expression of Emotion,” *J. Nonverbal Behav.*, vol. 34, no. 4, pp. 223–248, 2010.
- [26] E. Volkova, S. De La Rosa, H. H. Bulthoff, and B. Mohler, “The MPI emotional body expressions database for narrative scenarios,” *PLoS One*, vol. 9, no. 12, pp. 1–28, 2014.
- [27] F. L. Philippe, R. Koestner, S. Lecours, G. Beaulieu-Pelletier, and K. Bois, “The Role of Autobiographical Memory Networks in the Experience of Negative Emotions: How Our Remembered Past Elicits Our Current Feelings,” *Emotion*, vol. 11, no. 6, pp. 1279–1290, 2011.
- [28] T. Starner and A. Pentland, “Real-time American Sign Language recognition from video using hidden Markov models,” in *Proceedings of International Symposium on Computer Vision - ISCV*, 1995, pp. 265–270.
- [29] S. Shaily and V. Mangat, “The Hidden Markov Model and its application to Human Activity Recognition,” in *2015 2nd International Conference on Recent Advances in Engineering & Computational Sciences (RAECS)*, 2015, pp. 1–4.
- [30] I. Patel and Y. S. Rao, “Speech Recognition Using Hidden Markov Model with MFCC-Subband Technique,” in *2010 International Conference on Recent Trends in Information, Telecommunication and Computing*, 2010, pp. 168–172.

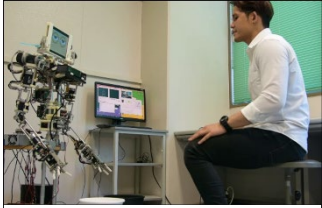
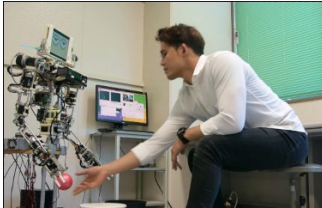
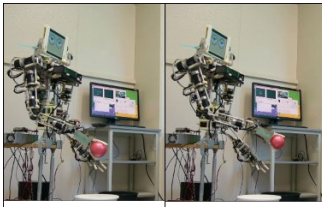
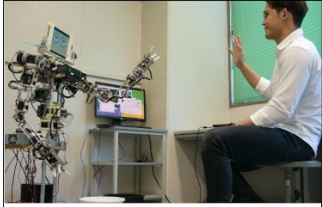
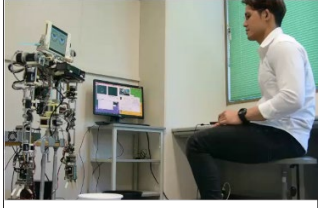
- [31] S. An, L. J. Ji, M. Marks, and Z. Zhang, “Two sides of emotion: Exploring positivity and negativity in six basic emotions across cultures,” *Front. Psychol.*, vol. 8, no. APR, pp. 1–14, 2017.
- [32] Y. Li and N. Zeng, “A Neural Network Based Inverse Kinematics Solution In Robotics,” in *Neural Networks in Robotics*, G. A. Bekey and K. Y. Goldberg, Eds. Boston, MA: Springer US, 1993, pp. 97–111.
- [33] E. Oyama and N. Y. Chong, “Inverse Kinematics Learning by Modular Architecture Neural Networks with Performance Prediction Networks,” 2001.
- [34] R. Kirby, J. Forlizzi, and R. Simmons, “Affective social robots,” *Rob. Auton. Syst.*, vol. 58, no. 3, pp. 322–332, 2010.
- [35] C. D. Kidd and C. Breazeal, “Effect of a robot on user perceptions,” *2004 IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IEEE Cat. No.04CH37566)*, vol. 4, pp. 3559–3564, 2004.
- [36] T. Nakata, T. Sato, and T. Mori, “Expression of emotion and intention by robot body movement,” *5th Conf. Intell. Auton. Syst.*, pp. 352–359, 1998.
- [37] T. Nakata, T. Mori, and T. Sato, “Quantitative Analysis of Impression of Robot Bodily Expression based on Laban Movement Theory.,” *J. Robot. Soc. Japan*, vol. 19, no. 2, pp. 252–259, 2001.



## Appendix

### Appendix A: List of predefined command pairing with robot tasks

Table A-1 Predefined command and robot tasks

Command	Words replying from robot	Robot task
Hello	Hello, Nice to meet you	 <p style="text-align: center;">Standby</p>
Take this	Yes, sir	 <p style="text-align: center;">Grip ball</p>
Game start	Ok, Let have fun	 <p style="text-align: center;">Hand over Box A      Hand over Box B</p>
Bye bye	Bye bye, see you again	 <p style="text-align: center;">Bye</p>
Set home	Yes, sir	 <p style="text-align: center;">Home position</p>

## Appendix B: Critical values of the $t$ distribution

Table B-1  $t$  distribution table

Cum. prob	$t_{0.50}$	$t_{0.75}$	$t_{0.80}$	$t_{0.85}$	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$	$t_{0.999}$
One-tail	0.5	0.25	0.2	0.15	.1	0.05	0.025	0.01	0.005	0.001
Two-tails	1	0.5	0.4	0.3	0.2	0.1	0.05	0.02	0.01	0.002
<b>df</b>										
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327
3	0.000	0.765	0.978	1.256	1.638	2.353	3.182	4.541	5.841	10.215
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.869	3.355	4.501
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098
Z	0.000	0.064	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%
	<b>Confidence Level</b>									

## **Appendix C: Publications and Presentations from the Present Research Work**

1. J. Inthiam, E. Hayashi, W. Jitviriyaya, and A. Mowshowitz, “Development of an emotional expression platform based on LMA-shape and interactive evolution computation,” in *Proceedings - 2018 4th International Conference on Control, Automation and Robotics, ICCAR 2018*, 2018, pp. 11–16.

2. J. Inthiam, E. Hayashi, W. Jitviriyaya, and A. Mowshowitz, “Mood estimation for human-robot interaction based on facial and bodily expression using a Hidden Markov Model,” in *2019 IEEE/SICE International Symposium on System Integration (SII)*, 2019, pp. 352–356.

3. J. Inthiam, A. Mowshowitz, and E. Hayashi, “Mood Perception Model for Social Robot Based on Facial and Bodily Expression Using a Hidden Markov Model,” *J. Robot. Mechatronics*, vol.31 , no. 4, 2019.

## List of Figures

Figure 2-1 Schematic representation of forward and inverse kinematics.	6
Figure 2-2 Denavit-Hartenberg frame assignment	7
Figure 2-3 ANN architecture	9
Figure 2-4 Single node diagram	10
Figure 2-5 Neural network as a black box	11
Figure 2-6 Diagram of multi-layer NN	11
Figure 2-7 Back-propagate error signals	13
Figure 2-8 Convolution neural networks model	14
Figure 2-9 Pooling process	15
Figure 2-10 Illustration of one-point crossover and re-assembles to get two novel solutions	18
Figure 2-11 The Gaussian distribution	20
Figure 2-12 Markov model with three states.	30
Figure 2-13 HMM example structure	32
Figure 2-14 Graphical represent of Viterbi's calculation	36
Figure 3-1 Overview of Seamless robot system	38
Figure 3-2 Configuration of head module	39
Figure 3-3 Structure of robot arm	40
Figure 3-4 Dynamixel RX-24F	44
Figure 3-5 Motor driver	46
Figure 3-6 STM32F4 Discovery	46
Figure 3-7 Diagram of control system	47
Figure 3-8 KinectV2	47
Figure 3-9 Multi-frame data (RGB frame, Depth frame, and Infrared frame)	48
Figure 3-10 Skeleton tracking.	48

Figure 3-11 Emotional expression by human	50
Figure 3-12 Triangle shape of upper body	50
Figure 3-13 Data collecting desk	52
Figure 3-14 Emotional expression data	53
Figure 3-15 Left- to right-hands distance	54
Figure 3-16 Hand gesture speed	55
Figure 3-17 Elbow angle movement	55
Figure 3-18 States Extracted from narratives	57
Figure 3-19 Frequency of occurrence state	57
Figure 3-20 Bodily mood model	58
Figure 3-21 CNN model for face emotion recognition	59
Figure 3-22 Emotion reliable scores	60
Figure 3-23 Face emotion classes	61
Figure 3-24 HMM of Moods perception model	62
Figure 3-25 Frame assignment for Seamless arm	64
Figure 3-26 Kinematics solution based Neural Network	65
Figure 3-27 Gestures based on LMA model	66
Figure 3-28 Chromosome structure	67
Figure 3-29 Overall gestures system diagram	68
Figure 4-1 Evolution of fitness value	74
Figure 4-2 Mood analysed result	75
Figure 4-3 Comparison of accuracy result by using two different input	76
Figure 4-4 Comparison of accuracy result between mood model and human evaluation	77
Figure 4-5 Interactive system	81
Figure 4-6 Infographic of guessing game	82
Figure 4-7 Seamless robot and human performed Guessing game	82



## List of Tables

Table 2-1 Basic Genetic Algorithm	17
Table 2-2 Motion factors and Effort elements	25
Table 2-3 Shaping Dimensions and affinities	27
Table 2-4 The probabilities for tomorrow's weather based on today's weather	30
Table 2-5 Viterbi algorithm	33
Table 2-6 HMM parameters for weather prediction model	34
Table 3-1 Specification of neck roll joint motor	39
Table 3-2 Specification of neck yaw joint motor	40
Table 3-3 Specification of shoulder joint motor	41
Table 3-4 Specification of elbow joint motor	42
Table 3-5 Specification of wrist joint motor	43
Table 3-6 Specification of hand motor	44
Table 3-7 Specification of motor driver	45
Table 3-8 Specification of STM32F4 Discovery	46
Table 3-9 Triangle parameters	51
Table 3-10 Lists of the states and their corresponding expressions	56
Table 3-11 ARCHITECTURE OF CNN MODEL	59
Table 3-12 HMM parameter for Mood perception model	62
Table 3-13 DH parameter of Seamless Robot left arm	64
Table 3-14 Procedure of human-interactive	69
Table 4-1 Initial gestures	71
Table 4-2 Final generation of positive expression	72
Table 4-3 Final generation of negative expression	73
Table 4-4 Paired t-test calculation from Positive clip videos	79

Table A-1 Predefined command and robot tasks	92
Table B-1 $t$ distribution table	93



## **Acknowledgement**

I want to express my special gratefulness to my advisor, Professor Dr. Eiji HAYASHI. I have learned many things as being your student. Thank you for allowing me to grow as a research scientist.

In particular, I am grateful to Professor Dr. Abbe Mowshowitz for enlightening me the first glance of research. Also, I appreciate your advisement and warm encouragement.

I also thank the 100th Anniversary Memorial scholarship. Your generosity allows me to take my goals and dreams a reality.

My thank also goes to my fellow lab-mates in HAYASHI LAB, for being friendly, for the sleepless nights we were working together before deadlines, and for all the fun we have had together. Thank you to everyone, who not mentioned here.

Last but not least, no words can express how I feel for what you have done for me so far. Special thanks to my family (Inthiam) for your encouragement. Without your patiently support, I could not become who I am today. Love you to the moon and back.

JIRAPHAN INTHIAM