ARMY HAND SIGNAL RECOGNITION SYSTEM USING SMARTWATCH SENSORS

by

Weonji Choi

Bachelors of Civil and Environmental Engineering, Korea Military Academy, 2013

Submitted to the Graduate Faculty of

Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This thesis was presented

by

Weonji Choi

It was defended on

May 29th, 2018

and approved by

Wei Gao, Ph.D., Associate Professor, Department of Electrical and Computer Engineering,

Murat Akcakaya, Ph.D., Associate Professor, Department of Electrical and Computer Engineering,

Zhi-hong Mao, Ph.D., Associate Professor, Department of Electrical and Computer Engineering,

Thesis Advisor: Wei Gao, Ph.D., Associate Professor, Electrical and Computer Engineering

ARMY HAND SIGNAL RECOGNITION SYSTEM

USING SMARTWATCH SENSORS

Weonji Choi, M.S.

University of Pittsburgh, 2018

The organized armies of the world all have their own hand signal systems to deliver commands and messages between combatants during operations such as search, reconnaissance, and infiltration. For instance, to command a troop to stop, a commander would lift his/her fist next to the his/her face height. When the operation is carried out by a small unit, the hand signal system plays a very important role. However, obviously, there is an aspect of limitation in this method; each signal should be relayed by individuals, which while waiting attentively for a signal can cause soldiers to lose attention on the front observation and be distracted. Another limitation is, it takes a certain period to convey signals from the first person to the last person. While the limitations above are related to a short moment, that can be fatal in the field of battle.

Gesture recognition has emerged as a very important and effective way for interaction between human and computer (HCI). An application of inertial measurement unit (IMU) sensor data from smart devices has lead gesture recognition into the next level, because it means people don’t need to rely on any external equipment, such as a camera to read movements. Especially wearable devices can be more adequate for gesture recognition than hand-held devices because of its distinguished strengths. If soldiers can deliver signals using an off-the-shelf smartwatch, without additional training, it can resolve many drawbacks of the current hand signal system.

In the battlefield, cameras to record combatants’ movement for image processing cannot be installed nor utilized, and there are countless obstacles, such as tree branches, trunks, or valleys that hinder the camera to observe movements of the combatants. Because of unique characteristics of battlefield, a gesture recognition system using a smartwatch can be the most

appropriate solution for making troops mobility more efficient and secure. For the system to be used successfully in combat zone, the system requires high precision and prompt processing; although accuracy and operating speed are inversely proportional in most of cases.

This paper will present a gesture recognition tool for army hand signals with high accuracy and fast processing speed. It is expected that the army hand signal recognition system (AHSR) will assist small units to carry-out their maneuver with higher efficiency.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I owe many thanks to many people for the support, help and kindness that they showed to me during the time that I strived for as a master student. I am very pleased to be able to express my gratitude in this page.

First of all, I am grateful to my family who are the reason and guide of my life. If there was no consistent and stable support they sent me, I would have already lost my way. Especially my husband Chanwoo, I believe that I do not need many words for you. Thank you and love you.

Thank you for the Republic of Korea Army who sent me to the United States and provided financial support for my study with faith. Professor Kim, Ho-gil, Oh, Gyeong-doo, and Yoo, Sam-hyeon. I will not forget the support you gave me for my new journey.

I am really grateful to my advisor, professor Gao, Wei, for your support. It was fortunate to be able to meet you. I look forward to having more chances to work with you again someday. And dear committee members, Mao, Zhi-hong and Akcakaya, Murat, thank you for helping me improve my study.

Finally, I appreciate my friends I met in Pittsburgh. You have not let me forget what is really important in my life. Because you were with me, I can go back home with many wonderful memories. My roommate Seoyoon, you have provided a space where I can get perfect relief and peace and you were the one who kept me sane.

Thank you for all who have been praying for me and Pittsburgh for being a lovely city.

1.0 INTRODUCTION

In this thesis I demonstrate how a gesture recognition system read army combatants' hand signal. The system is an optimum solution for current army hand signal protocol, because Army Hand Signal Recognition (AHSR) system is a robust real-time system that requires only a wrist-worn wearable device and doesn't require any further training nor external equipment. This chapter contains a motivation for this study, my own contributions, and a description of the structure of the entire paper.

After World War II, the aspect of modern warfare (aerial, intelligence, network-centric, and nuclear etc. [1]) appeared accompanied by the evolution of technology; however, small unit maneuvers are still considered as a fundamental aspect of warfare and the significance should not be underestimated. During a maneuver operation, combatants rely on agreed communication method among themselves which can be voice, radio, or hand signal. The hand signal is needed when sound should be limited because the unit is too close to the enemy or the distance of individual each soldier is not close enough. The hand signal, which is necessary to some battlefield environments, shows some inevitable limitations; each signal should be relayed by individuals, which while focusing attentively for a signal can cause some soldiers to lose attention on the front observation and be distracted. Another limitation is, it takes a certain period of time to convey signals from the first person to the last person. Lastly, the signals itself can be missed in the process of being conveyed; in other words, the successful conveyance of command to the last soldier may not be guaranteed. While the limitations above are related to a short moment, that can be fatal in the field of battle.

During recent years, within computer engineering, pattern recognition has made remarkable growth. The pattern recognition refers to the automatic detection of regularities in input data by using algorithms, and the conduction of further works such as categorizing the data into different classes with the use of former regularities [2]. Gesturing is a means of communication that people feel very natural and do in their daily life. If the gestures can be used to control something using pattern recognition, such as a system or movement of the object, it is very efficient and its application range is and unmeasurable wide way for human-computer interaction (HCI). Vision based pattern recognition has realized many imaginative things in our life, for instance air mouse, virtual personal trainer [3]–[6]. The vision based system is very powerful; however, it also has inevitable limitations and thus its availability is restricted under some conditions. The vision based system is not tolerant to environment, such as illumination, surrounding objects, the orientation of camera and subject. One method to overcome this limitation is to use the sensing data of the device on which sensors are mounted; a typical example is the Wii remote controller.

Nowadays, the smartwatch is one of the most popular personal devices along with the smartphone. It is as light as a usual analog watch and relatively cheap, but it provides a wealth of functionality (such as microphone, speaker, camera, and IMUs) and does not restrict, force, or change a user’s movement. Further, the smartwatch worn on user’s wrist is always ready to be used and to communicate wirelessly with other devices [7], but user does not have to touch the device or even need to look at it. These features imply that the smartwatch is very suitable equipment to be used for analyzing human hand or arm movement.

The combination of pattern recognition and wearable device technology produces a boundless synergy. How can the strength of each technology be integrated for robust and accurate gesture recognition system? Pattern recognition using wearable device has been primarily used for fitness, sport medicine, rehabilitation, health monitoring, and so on. But this paper shows another application of pattern recognition for military purpose. Chapter 2 describes the growth of pattern recognition in more detail and highlights studies already done by other researchers.

Of course, there were a few restrictions and obstacles in developing a gesture recognition system to be used in the battlefield. First, gesture recognition is a different concept than activity recognition supported by most current smart devices. Activity recognition is aimed at recognizing a specific activity by detecting the repeated movement or status change over a certain period of time. For example, Android provides five types of APIs for activity recognition; driving, bicycling, running, still, and walking[1]. On the other hand, a gesture recognition system has a difficulty in detecting gestures that are performed in a relatively short period and may not be repeated.

Second, there are several difficulties in converting signals into identifiable input gesture. Signaling movements are 3 dimensional gestures, not 2 dimensional, and each soldier has different ways of drawing the same signal. For instance, the angle of his/her arm or the path to reach one point to another will be slightly different from soldier to soldier. One of the well-known gesture control application is a music player control system which only requires four kinds of gesture to play, stop, forward, and backward. On the other hand, an army hand signal

[1] https://developer.android.com/training/gestures/detector.html?q=activity%20recognition

system has relatively higher number of gesture types. And some gestures are not distinguishable enough to find differences at a glance.

Lastly, extra ordinary characteristics of a battle field produce more restraints. There is no guarantee of reliable network or pre-built technological infrastructure. These environmental aspects imply that the system to be used in the field must have both a certain level of accuracy and speed; although accuracy and operating speed are inversely proportional in the majority of cases.

It is hoped that this research serves as a foundation for future development of a system which can recognize combatant's every single movements of hands and fingers. This study concentrates on 14 kinds of one-hand signal for maneuver operation. The reason this start with maneuver operation first is because, during a maneuver, each combatant is likely to be far away from each other and close to an enemy area where sound is not suitable for communication. Based on the experimental results in this paper, the system shows the high possibility of expansion to be applied to further various gestures.

1.1 MOTIVATION

Korea is the only divided nation in the world. The Korea peninsula is divided by a Demilitarized Zone (DMZ) that is 248km long and traverses the entire peninsula 4km along the Military Demarcation Line (MDL) which is near by 38th parallel. Many soldiers carry out search and reconnaissance operations every night in the DMZ. Further, Korea has more mountainous regions than flat regions and the DMZ is no exception. When a small unit maneuvers under these circumstances, various obstacles, like tree branches, trunks, bushes and valleys, block the view of each combatant. This environment makes it difficult for combatants to use hand signals. It is even more unreasonable to use voice or radio to communicate in the DMZ where enemy forces and friendly forces are coexisting so closely. Careless sounds can expose the unit's position and make it possible to be in the enemy's surveillance area. The current alternative is to use hand signals or connect a headset or earphone to the radio. However, the biggest blind spot of this resolution is that it also cuts off valuable sounds which may come from the surrounding area, such as the enemy's approaching sound. So, in order to make this precarious situation better, I suggest AHSR that can improve the current hand signal system.

Important initial condition of the system was that it should not require any additional training nor external equipment. In modern warfare, combat can occur everywhere, so the system must be independent of the environment. Further training means a lot of potential confusion and time to adjust to the new method. Also, any additional body attachment equipment may reduce the soldiers' combat performance. Therefore, the hand signal recognition system using smartwatch sensors can be an optimum answer to improve the current analog hand signal system. It not only solves the limitations of current hand signals but also has a wide range of utilization. The accumulated data retrieved from the AHSR system can be used for operation performance analysis after completion of the operation and for educational purposes through the use of case studies. Furthermore, soldiers can customize the system to add and train their own gestures, so that a small unit can make the team's unique hand signal protocol. While presenting these strengths of the system, the smartwatch has its own benefits. Due to its rich functionality, the smartwatch is capable of a variety of tactical uses, not just for AHSR; it can be used as a map viewer, walkie-talkie, or statement delivery system.

1.2 CONTRIBUTION

As mentioned above, several challenges surely appeared, however, in other words, those challenges lead to a contribution of the research in this paper. First, this paper suggests a new application of the gesture recognition. Many studies have caused numerous fields and areas, but not for military purpose. The computer-human interaction using gesture is a very important stream in the field of computer engineering in modern times. As military also has several areas where this trend can affect, this study is expected to be the starting point.

The system in this paper used smartwatch which is considered as one of the most suitable equipment for gesture recognition at present. Due to the strength of the smartwatch, a soldier does not need to change the movements of his/her hands and arms in order to use the system. In other words, the system is ready to be used without any additional training.

This study compared and evaluated many factors for recognizer design using a relatively large number of gestures; algorithm, variables, features and required optimal number of training data. This large amount of empirical experiment results guarantees the scalability of the system.

Most of all, the accuracy and the operation speed of the system is high and fast enough to be used for tactical purpose. The accuracy was over 99% and the time delay for gesture recognition was less than 1 sec with entirely based on real-time operation.

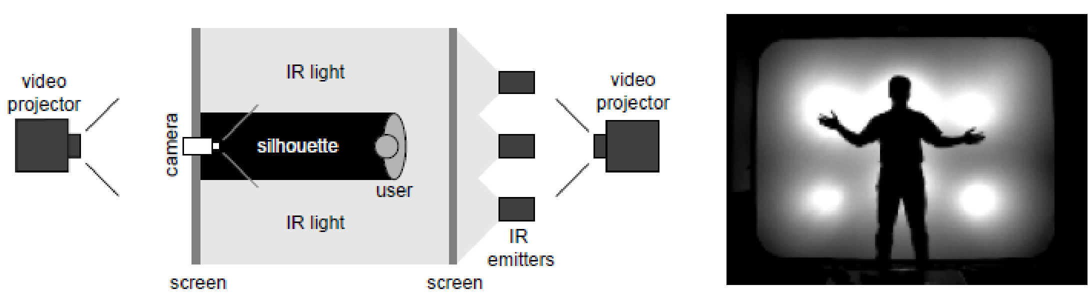
1.3 THESIS ORGANIZATION

In chapter 2, literature review will be introduced and in chapter 3, the theoretical background for this study, which will help to understand the following content, will follow. Chapter 4, as a main part of this paper, will explain every detail of experimental process, data collection, optimal modeling, and the Army Hand Signal Recognition system (AHSR). Finally, the last chapter will discuss limitations of this study and suggest potential prospective work.

2.0 RELATED WORK

During recent years, within computer engineering, pattern recognition has made remarkable growth. As pattern recognition fused with the gesture, gesture recognition, it's application range became very broad. Because of the development of smart devices, gesture recognition systems had become an important way to facilitate human-computer interaction (HCI). If the system can utilize sensor data from the smart device, it means people don't need to rely on any external equipment, such as a camera to read movements or illumination to maintain constant brightness. Furthermore, gesture recognition has grown to a higher level along with the development of wearable devices. Chapter 2 follows the flow of the gesture recognition's advancement. First is the previous work of the early stage of gesture recognition which was mostly vision based. Second is the convergence of the sensor data of hand-held smart devices and gesture recognition. Last is the recent skyrocketed interest in gesture recognition using the wearable smart devices.

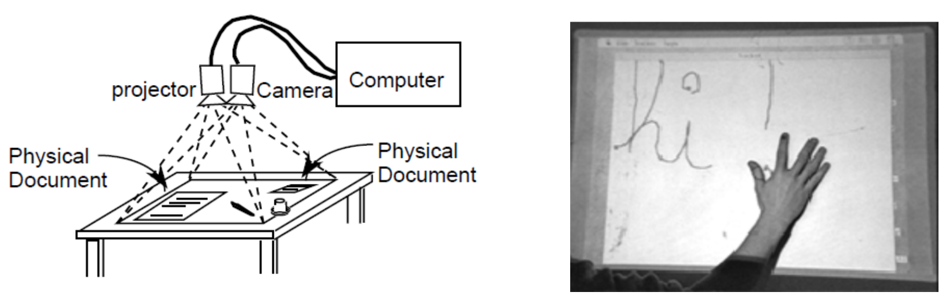
During the early stage of gesture recognition, vision based recognizing systems ([3]-[6])



(a) (b)

Figure 1 (a) A conceptual drawing for Virtual PAT; (b) A image filtered through a visible-block filter [4].

were main stream and also can be categorized as a computer vision technology. Basically, several pieces of external equipment are required for a vision based system, such as a camera, projector, or illumination to guarantee minimum brightness. James W. Davis et al. presented a very practical vision based gesture recognition system in 1998, a virtual personal aerobics trainer (Virtual PAT) using a camera, video projector, IR light, IR emitters [4]. Figure 1a shows a conceptual drawing for Virtual PAT; a camera of the PAT system records images of a room and the subtracts silhouette of a user using optical blocking of specialized non-visible light, e.g. infrared light (Figure 1b). After the subtracted silhouette is compared with the system's motion templates, if each aerobic movement is complete, positive comments (e.g. "good job!", "fantastic!") are provided, and if not, negative feedback (e.g. "get moving!", "concentrate!") are provided. James L. Crowley et al. introduced the digital desk utilizing a projector and video camera for finger tracking (Figure 2a) [5]. A computer screen is projected onto a physical desk using the projector, and the camera is set up to observe the work area. The system tracks some pointing devices, such as a finger, a pencil or an eraser, to determine the most likely location of the object at some moment. The authors suggested an application of their system, "finger-paint" (Figure 2b). The vision based system has suggested a new way for HCI, the 'gesture', which refers to means of communication that people feel very natural and do in their daily life, however,



(a) (b)

Figure 2 (a) The Digital Desk for finger tracking; (b) Drawing and placing with "finger-paint" [5].

it still has some limitation. The vision based system is not tolerant to environments, with illumination issues, surrounding objects, varying orientation of camera and subject. Therefore, it can be very powerful only when it can be operated in a fully equipped space and with well-prepared conditions

Later, the use of sensor (e.g. IMU) data from the hand-held device has greatly expanded the scope of application for gesture recognition. A typical example of the hand-held device, a smartphone is now an essential device to most people, and has a variety of built-in sensors, e.g. magnetometer, proximity sensor, barometer, gyroscope, and accelerometer. Use of a hand-held device solves the vision based gesture recognition's limitation, which has a high dependency on external conditions [8]–[11]. In 2013, Sven Kratz et al. suggested a combining acceleration and gyroscope data for motion gesture recognition in their paper [8]. If the recognizer only uses acceleration data without gyroscope data, rotation data should be approximated using accelerometer and no tilt and rotation information can be obtained when the device is rotating on the plane perpendicular to gravity. Their hypothesis was to prove that the system can recognize more complex gestures with higher accuracy by using a combination of accelerometer and gyroscope data. They proposed detailed results with a comparison between different combinations of sensors and classifiers, by using data from accumulated experiments. As a result, the authors concluded that the combination of acceleration and gyroscope data can improve the gesture recognition rate up to 4% using 3 classifiers, Protractor3D, DTW (Dynamic Time Warping), and LR (Regularized Logistic Regression). Michael Hoffman et al. presented a systematic study on the recognition 3D gestures using Nintendo Wii Remote for the linear acceleration-sensing and Nintendo Wii Motion Plus for the angular velocity-sensing (Figure 3) [9]. In the paper, the experiments proceed to find how many gestures can each classifier recognize, how many training samples are needed per gesture, proper classifiers to achieve some degree of accuracy, and also to assess user dependency of a recognizer. As a result, their developed system could recognize and classify simple gestures that can be used for gaming purposes with 99% of accuracy using linear and AdaBoost classifiers. The use of hand-held devices provided a significant improvement of the gesture recognition field; with this, the system became environmentally independent, so there were little spatial constraints. However, the hand-held device still showed obvious restraints; the user had to continue to hold or grab a device, which meant that the results can vary significantly on the angle at which the user was holding the device. Furthermore, the hand-held device restrained or changed the user's original movements.

Nowadays, the smartwatch is one of the most popular personal devices and can be used as an optimal device for gesture recognition due to its strengths; it is relatively cheap, providing a wealth of functionality, which is always ready to be used and communicate wirelessly with other devices. Furthermore, the wearable device provides a higher level of freedom to a user's movements than the hand-held device. In other words, a smartwatch does not restrict, force, or change a user's movements. There are a large number of research for gesture recognition using wearable devices, which are very practical and have distinct focuses; gaming [12], health care [13], [14], sports medicine [15], rehabilitation [16], interaction with other devices [7], [17]–[19], and handicapped aids [20]. There are also studies to improve the performance of gesture

Figure 3. A user providing 3D gesture data using a Nintendo Wii hand-held device for [9]

recognition itself [21]–[25]. Furthermore, it is possible to use both the hand-held device and the wearable device together for a single purpose system [26], [27]. Keiko Katsuragawa et al. described the Watchpoint, a hands-free, smartwatch-based and mid-air pointing interaction system [7]. Their work demonstrated that an off-the-shelf smartwatch, not any other vision based or specialized device, can be used as an interaction device for pointing in ubiquitous display environments. Because the Watchpoint doesn't require a vision-based tracking system, the occlusion problem, which is the major limitation of vision based gesture recognition systems, can be solved. It means that the user and display can interact even when there is an obstacle between them. The paper written by Lorenzo Porzi et al. also shows an example of gesture recognition's practical application designed for people with visual impairments [20]. The user can make a gesture as an input to make the system recognize the gesture and activate a corresponding function. Their research demonstrates that the interaction between visually impaired people and wearable devices can be possible by utilizing gesture recognition.

The current research done on gesture recognition has caused many fields and areas. The gesture recognition suggested a way of human-computer interaction that can be used broadly for the general public. However, no attention has given to the application of gesture recognition for military purpose. With this regard, this paper suggests a very practical utilization of gesture recognition for military fields.

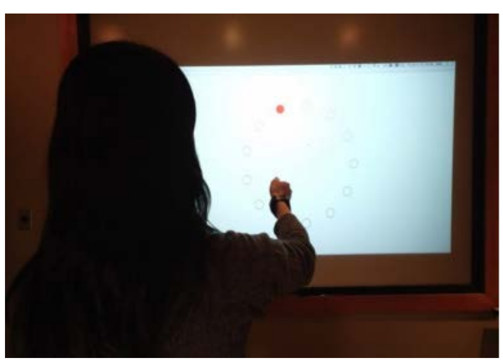


Figure 4. A photo of demonstration of [7].

3.0 THEORETICAL BACKGROUND

3.1 PATTERN RECOGNITION

The pattern recognition is related to the automatic detection of regularities in input data by using algorithms, with the use of these regularities, to do further work such as categorizing the data into different classes. One of the practical uses of pattern recognition is a handwriting recognizer (for letters and digits) used in post offices, banks, and smartphone apps [28]. The well-known pattern recognition theories are Bayesian Decision Theory, Probability Distribution, Dynamic Time Warping (DTW), Neural Networks, Support Vector Machines (SVM), Hidden Markov Models (HMM).

In most case of practical use, the original input data or variables should be preprocessed to convert them into some new space of variables to make the pattern easier and faster to be recognized. This pre-processing stage can also be called feature extraction. In the handwriting

(a) (b)

Figure 5. Example of input for pattern recognition's practical use for (a) handwriting recognition, (b) army hand signal recognition.

recognizer example above, the images of the writing (Figure 5a) are typically interpreted and scaled so that each digit can be contained within a box of a fixed size [2]. For AHSR, the human hand's movements (Figure 5b) are translated and digitized to the acceleration and angular velocity using sensors of smartwatch.

For the pattern recognition, there are two kinds of data sets, training data and test data; the training data set is a group of samples that will be used to build recognizer, and the test data set is a group of samples that are used to evaluate the performance after creating the recognizer. Both the training data and the test data should be preprocessed in the same way [28].

The sensor data representing the hand's movement for AHSR corresponds to the time-continuous data like a seismic wave, voice, and the daily values of a currency exchange rate. These lists of data are usually called as sequential data. When analyzing sequential data, the temporal relationship must be considered significantly; if the order of the features change, the physical characteristics of the sequential data's pattern can be distorted. Another important characteristic of sequential data is that each data has a different time length. The movements for the hand signal are performed at different speeds, therefore, each has a diverse time length for each person and even each trial. Among the pattern recognition theories mentioned above, most suitable theories for the sequential data would be Dynamic Time Warping (DTW) and Hidden Markov Models (HMM). While these two techniques share (partially) mutual concepts, the operations and the results were totally different. B.-H. Juang explains in-depth in his paper about the similarities and differences between the two algorithms in terms of speech recognition [29]. The following sections will address the theoretical basis of the two theories to support understanding of the remaining paper.

3.1.1 Dynamic Time Warping

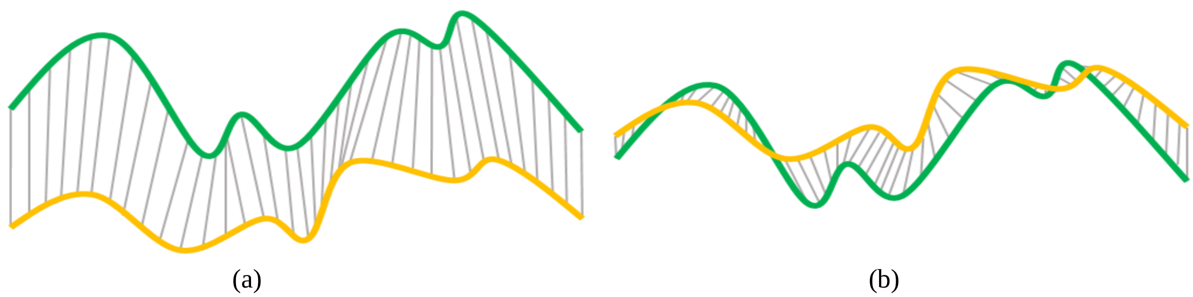


Figure 6. A graphical description of Dynamic Time Warping (DTW). (a) displays two different input signals desired to be classified to same class; (b) shows one of the candidates for the best match of two signals.

In order to analyze sequential data and recognize a pattern, a model that can appropriately express the temporal property contained in the data and deduce the desired information from the data is required. Dynamic Time Warping is one of the most popular algorithm with HMM for measuring similarities between two different sequence data which can have different time length or execution speed by warping them and calculating the costs to find the best match of them. Although DTW has been generally used for speech recognition [30][31], it can be applied to any pattern recognition problem of continuous data [8], [17], [20], [25], [32]. As mentioned above, the sequential data has its own characteristics; the order of features need to be considered most importantly and each data's time length can be varied. Especially, hand signals, input data for AHSR, are relatively lengthy and compound movements; it implies that a lot of efforts can be required for preprocessing. DTW can solve these difficulties effectively. Figure 6a shows two sequential data that my look similar but not exactly same; the number of peaks and dips are identical but the distance between each bump are different. Grey lines are marked to enhance understanding about how DTW determines the best match. Figure 6b illustrates DTW calculated the sum of the length of the grey lines and the findings about the minimum value, which indicates optimal.

The trend of transition from DTW to HMM occurred in the late 1980s because DTW is no longer powerful enough for stochastic signals that the large number of real-world data corresponds to. For alternative way for standard DTW, stochastic DTW was designed, which is closely related to HMM that will be introduced in the next section [30]. HMM-based methods show better performance when the training data is sufficient, however, DTW is still considered very effective especially when the number of training data is limited.

3.1.2 Hidden Markov Model

A hidden Markov Model (HMM) is the most representative and widely used model for sequential data. In general, HMM is a type of stochastic modeling suitable for nonstationary stochastic sequences whose statistical properties undergo distinct random transitions among a set of k different stationary processes [33]. HMM is a combination of ‘hidden’ and ‘Markov model’. Markov model uses first order Markov chain (Figure 7) invented by Andrey Markov and has been used for various applications. However, some limitations emerged when dealing with more complicate procedure, so the need of development of HMM was raised. If the model uses Markov chain, its size increases exponentially according to the order; the higher order could be

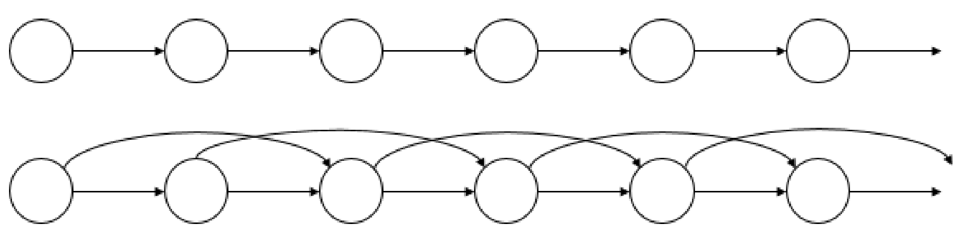


Figure 7. 1st (above) and 2nd (below) order Markov chain

an indication of more accurate modeling but also realistically uncontrollable number of parameters.

However, the order doesn't need to be fixed for HMM, and the model will determine adaptively in the process. HMM is using 'hidden states' to make the model's size affordable. Figure 8 is a transition diagram explaining the basic concept and showing two representative types of HMM; Figure 8a is called ergodic model and 8b is called left-right model. The ergodic model is used to express fully connected HMM, and left-to-right model is known to be suitable for speech recognition and can be extended to parallel left-to-right model. Determining which model should be used for a recognizer design can be done empirically.

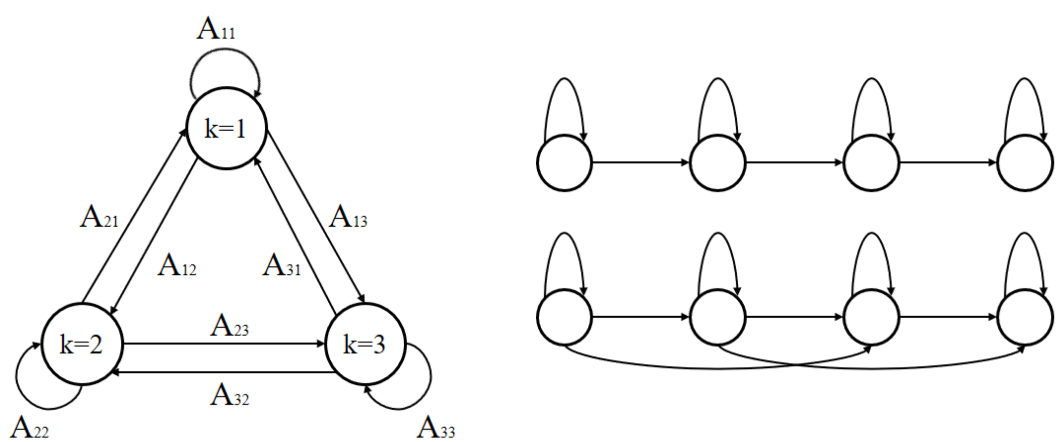


Figure 8. Transition diagram showing a model and architecture of HMM for (a) a 3-state ergodic model. A_{ij} denotes the elements of the transition matrix; (b) a 4-state left-to-right model.

HMM can be defined as $\lambda = (A, B, \pi)$, when A is the transition matrix, B is the observation probability matrix and π is the vector of the initial probabilities. There are three basic problems for HMM - evaluation, recognition, and training. First, when the model is given, HMM will calculate the probability of the observation sequence $O = O_1 O_2 \cdots O_T$; the evaluation is used for recognizing some applications. For this study, the evaluation was used for the gesture classification. Second, the recognition is a problem that HMM to find optimal corresponding state sequence $Q = q_1 q_2 \cdots q_T$ (i.e., the Q can explain the observations in the best way). Lastly, the training is process adjusting parameters of the model to maximize the probability of the observation, $P(O|\lambda)$, using given observation sequence O. Each problem has prominent algorithm; forward-backward algorithm for evaluation, Viterbi algorithm for recognition, and Baum-Welch algorithm for training problem. This study also used these most well-known algorithms.

In the beginning, HMM was also used for speech recognition [34], [35], but the range of its application field has been gradually expanded; for gesture recognition [12], [21], [36]–[38], computer vision, data mining, bioinformatics [39] and etc.

4.0 ARMY HAND SIGNAL RECOGNITION SYSTEM (AHSR)

Chapter 4 is a main part of this paper. First section displays detailed description of experimental setup including explanations of for 14 gestures from current hand signal system of Republic of Korea Army (ROKA), used device, data transmission method, software and comprehensive experimental procedure. Second section describes data collection with information of participants, data collection process and the results. The section 4.3 is for optimal modeling; the section introduces variables that the designer can adjust to build more adequate recognizer, result comparison from different variables and algorithms, comprehensive evaluation, and conclusions. The Army Hand Signal Recognition system design based on former sections' result will be introduced and evaluated in the last section, which will guide to the conclusions.

4.1 EXPERIMENTAL SETUP

4.1.1 Army Hand Signal System

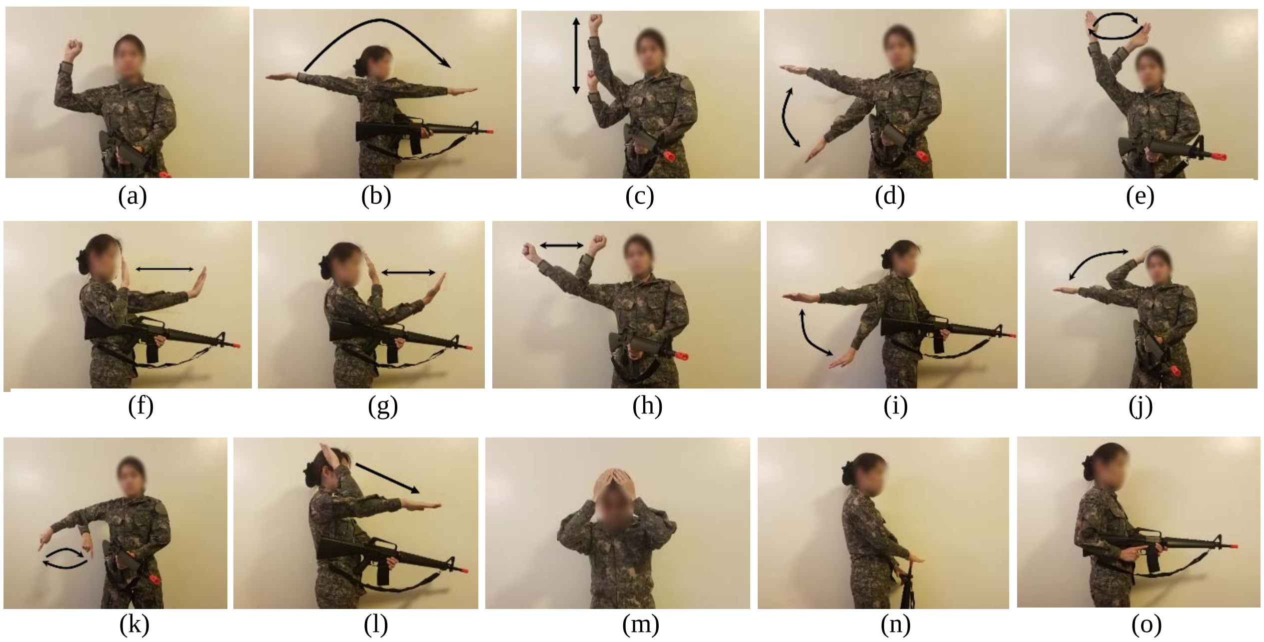


Figure 9. Pictures of ROKA's hand signals for maneuver operations

ROKA use many kinds of hand signal, and sometimes each unit make its own signal as a method of communication. For this paper, 14 gestures that are considered as the most general and widely used by entire ROKA for maneuver operations were selected to train and test AHSR. The signals can be divided into posture(pose) and gesture(movement) in more detail. If a stopped action means a command, then the signal will correspond to a posture. On the other hand, if a sequential movement means a command, the signal will be categorized as gesture. Figure 9 contains pictures for each 14 gestures ((a) – (n)) and one basic pose (o); the first three pictures correspond to posture and the rest of them correspond to gesture. For detailed description of each gesture, refer to Appendix A; each command description of each gesture was eliminated due to security purpose and distinct ID number was imposed instead. A M16A1[2] replica toy rifle was used for this study to help users perform gestures as real as possible and each gesture took 2.5 to 3.5 seconds in average to be performed. Figure 9o shows the posture of holding a rifle when soldiers are not aiming at a point. Therefore, in this study, the participants were required to take the default posture all the time except making hand signals.

4.1.2 Inertial Measurement Unit Sensors of Smartwatch

(a) **(b)**

Figure 10. (a) Official picture of LG Watch Urbane (LG-W150); (b) Urbane worn on user's wrist.

The Inertial Measurement Unit (IMU) is an electronic device that measures an object's specific force, angular rate, and magnetic field using an accelerometer, gyroscope, magnetometer, or combination of these. IMUs are widely used as part of navigation equipment for the aircrafts, satellites, and unmanned aerial vehicles (UAVs). All experiments for AHSR were conducted with android smartwatch LG Watch Urbane (Figure 10). Urbane is equipped with IMUs supported by 9-Axis accelerometer, gyroscope, and compass; Urbane also has PPG (Heart rate

[2] M16A1 rifle was developed in 1957 and adopted as a gun for the US Army in 1967. It is a representative rifle used in the Vietnam War and adopted as a rifle in Korea and other countries besides the United States. The length of M16A1 is 99cm, similar to the length of k-2, mainly used by Korea Army.

monitor), Barometer.[3] Using these hardware-based sensors, more than 10 kinds of software-based sensors deriving data by mimicking hardware-based sensors are available, such as the linear acceleration sensor or the gravity sensor[4]. The hardware-based sensors used for this study are the accelerometer and the gyroscope, most widely used by IMU sensors. The detailed specifications of the accelerometer and the gyroscope are shown on Table 1 (these can be obtained by using SensorList API). Sensor API allows users the users the right to set the data delay; the data delay controls the interval between transmissions of sensor data. Four basic options (Table 2) are provided, and the user can also specify the delay as an absolute value (in microseconds) as of Android 3.0 (API Level 11). For this study, time delay was established for the game mode (20,000 microseconds), the empirically determined speed which is fast enough to measure changes of the values.

Table 1. Detailed specification of the accelerometer and the gyroscope

Sensor Name	Vendor	Maximum Range	Power	Minimum Delay
MPU6515 Accelerometer	InvenSense	19.613297	0.4	5000
MPU6515 Gyroscope	InvenSense	34.906586	3.2	5000

Table 2. The time delay options provided by Sensor API

Delay type	Time delay (microsecond)
SENSOR_DELAY_NORMAL	200,000
SENSOR_DELAY_GAME	20,000
SENSOR_DELAY_UI	60,000
SENSOR_DELAY_FASTEST	0

[3] http://www.lg.com/us/smart-watches/lg-W150-lg-watch-urbane#
[4] https://developer.android.com/guide/topics/sensors/sensors_overview.html

The accelerometer provides 3-axis acceleration force (m/s^2) including gravity. Therefore, when the user does not move his/her wrist, the accelerometer reads a magnitude of $g = 9.81\ m/s^2$. Figure 11 shows the coordinate system of the smartwatch. The gyroscope measures the rate of rotation around each 3-axis (rad/s). The rotation in the counter-clockwise direction indicates positive values. And one more software-based sensor was used for this study, the linear accelerometer. Its concept is similar to the accelerometer, however, it excludes gravity. Same coordinate system used by accelerometer is also used for linear acceleration and gyroscope data.

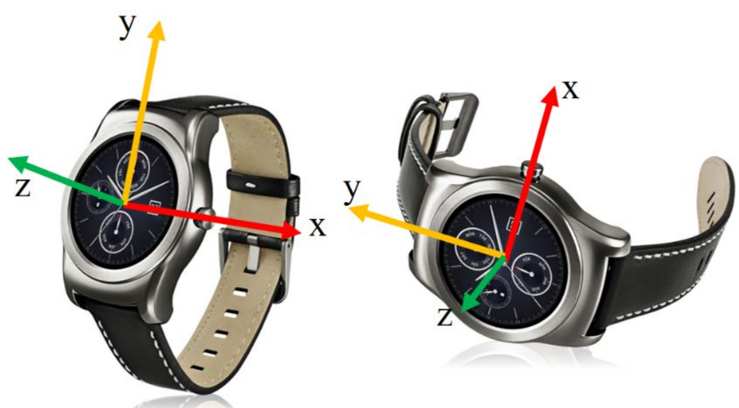


Figure 11. Coordinate system of smartwatch

4.1.3 Software

For this study, data was transferred in real time from the smartwatch to PC through the WIFI and saved as a .csv file with proper heading and format on the PC if needed. The application for smartwatch side used 'onSensorChanged' API to trigger the app to send the sensor data. A user can enter or change the IP address and port number of the server PC on the smartwatch screen. Also, the GUI support an on/off switch to start or stop the data transmission and text labels show the change in the sensor data value in real time; the user can check the operating status of the sensor using the GUI (Figure 12).

(a) (b)

Figure 12. A screenshot of smartwatch when (a) data transmission is off; (b) data transmission is on.

During the experiments for features and the algorithm selection, the part of Gesture Recognition Toolkit (GRT) designed by Nicholas Gillian and Joseph A. Paradiso of *Responsive Environments Group, Media Lab, Massachusetts Institute of Technology* was modified and used. The GRT is a cross-platform open-source C++ library that was developed to primarily make real-time machine learning and gesture recognition with emphasis on convenience and flexibility of users' customization; there are GRT wiki[5] and GitHub[6] for further information. The GRT supports various stages for gesture recognition, such as preprocessing, feature extraction, classification, regression and so on. For this study, adapted code of DTW and HMM were used [40].

When designing a model with HMM, several options were available for the engineer. First, there are two most popular architecture for HMM, ergodic and left-to-right model. The left-to-right model is often used for situations where a part of a pattern does not appear again

[5] http://www.nickgillian.com/wiki/pmwiki.php?n=GRT.GestureRecognitionToolkit

[6] https://github.com/nickgillian/grt/wiki

after it has passed, for instance, the speech recognition. Experiments were conducted to determine which architecture is more appropriate for the system, and the results are discussed in chapter 4.3. Furthermore, the GRT provides users with more freedom in recognizer design aspects; values of sigma, down-sample size, committee number. In order for the sigma parameter to determine a good match with the model, it controls the necessity of the proximity of each input vector to the model. The down-sample size measures how much each training data is down-sampled using average value to create each state in the model. If a user set the down-sample size as 10, the length of the input data will be changed into 1/10 of original length, and the data will be modified to calculated average values of each ten data. Lower down-sample size can speed up the training process, but if the value is too low, there is a risk of losing valuable data. Using the committee size, the designer can control the number of models that will be used to make a prediction. The experiments were operated to find optimal modeling with different combinations of these adjustable variables. The impact of each variable on this study and the optimal results will also be discussed in section 4.3.

A server-side software to record input data on PC and the analysis of the result were implemented with C++. When the user had chosen a gesture ID before recording training dataset, the ID has become a part of name of each data file. After the completion of the recording the training data for all types of gesture by the user, a training process for the recognizer can be finished by one line of command including a type of algorithm, the values of tunable variable, and a list of selected features.

4.1.4 Experimental Procedure

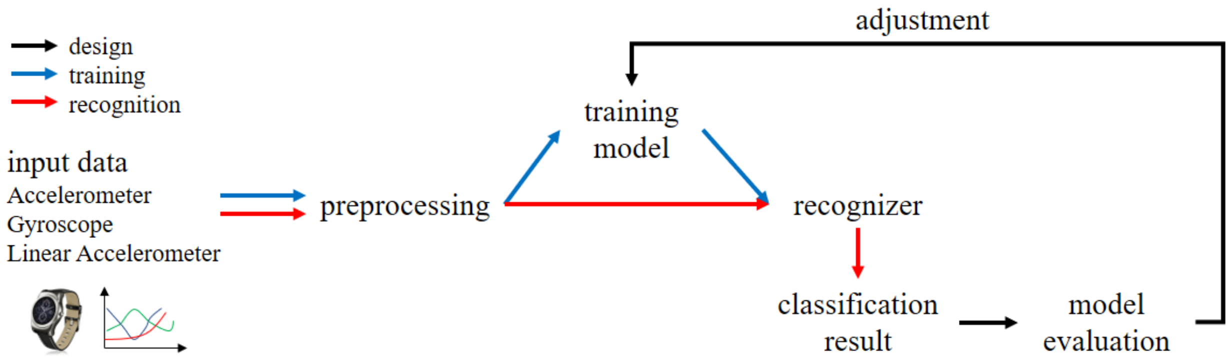


Figure 13. Conceptual Diagram for Designing Process

Figure 13 shows overall experiment procedure to find optimal recognizer design for AHSR. The blue line indicates training process, the red indicates recognition, and the black line refers to the designing process. After initial model is trained by input data, the test dataset evaluates the recognizer. Using the results from the first evaluation, the variables will be adjusted to improve recognition precision. The procedure shall be conducted repetitively to find an optimal design for the recognizer. Lastly, the recognizer with the optimal design would be verified with more participants' data.

4.2 DATA COLLECTION

4.2.1 Participants

For the feature and the algorithm selection procedure, the dataset from an army soldier trained for 9 years was used. The army soldier conducted each 14 types of gestures over 70 times to cumulate abundant dataset. To collect data for validation of the selected features and the algorithm, 10 participants were recruited. Each participant had to spend about 40 minutes and were compensated for their time and effort. They were asked to fill in a brief survey (refer Appendix B) about their demographic information, agreements for information provision, and the feedback for the system's sustainability. Summarized demographic information of participants are shown in Table 3 (Refer Appendix C for detailed information of each participant).

Table 3. Demographic summary of participants

Sex	**#**	**Mean age** (years)	**Mean height** (cm)	**Left-hand user**	**Smartwatch user**	**Military trained**	**Gesture recognition aware**
Male	4	34	174.03	-	1	3	-
Female	6	28.3	165.8	2	1	-	1

The participants used their primary using hand, and 10 participants in total were involved in the experiment. There were 8 right-handed and 2 left-handed users. Before starting to make gestures, the participants were given a brief description of the experiment's motivation, goal, and process, and an instruction was provided, including a table describing each hand gesture. Only 2 participants were smartwatch users, and most of the participants were non-military people and non-expert in gesture recognition field. The army hand signals system is very unfamiliar gestures to participants with no experience in military training, and it varied each time they performed the gesture. They are also unware of the sensitivity of the smartwatch sensor, the detailed classification principles of the recognizer, and the process. It indicates that it was difficult for the participants to make any artificial effort to produce better results. In other words, the results from the 10 participants are highly objective, as it is expected that higher accuracy will be obtained if a user is military-trained person and more education about the system can be done in advance.

4.2.2 Data Recording Procedure

Data was transferred in real time from the smartwatch to PC through the WIFI. The time delay for data transmission was 0.3 sec in average. The detailed programmatical part was introduced in section 4.1.3. This section focuses on the interaction between participants and the system. A footswitch[7] (Figure 14) was used in order to provide convenience to participants and indicate the starting and the ending point of each gesture during data recording procedure. The participants were able to choose whether to manipulate the switch by themselves or by a researcher on behalf.

Figure 14. A footswitch used for data recording

A fairly large amount of data was needed to test various options for recognizer modeling during the experiments; all gestures were repeated 70 times and it took about 5 hours in total. On the other hand, the all 10 participants were required to repeat only 16 times per gestures (13 for training data and 3 for test data) for the evaluation of the recognizer, and it took 40 minutes in

[7] www.pcsensor.com/usb-keyboard/one-switch-pedal-/usb-foot-switch.html

average; because the experiment for the optimal modeling proved that 13 training data can achieve sufficiently high accuracy. The participants had to abide by a rule not to adjust any circumstances of the watch including its position and angle, until they finish recording for at least one complete set of gesture. Because smartwatch is a relatively small device, it is quite sensitive to its position.

The participants finished recording one type of gesture and then passed on to the other kind of gesture. A single gesture file stored in '.csv' file named with each gesture's unique ID number and a group of .csv file for training data was then merged into one '.ahsr' file. Thus one '.ahsr' file was created per participant. To evaluate real-time AHSR system, each participant performed all 14 types of gesture in a random order, and the real-time recognition results were evaluated.

As a result, when calculating the pure time of gesture data, more than 160 minutes of data was accumulated.

4.3 OPTIMAL MODELING

This section contains the process of finding the best conditions for optimal modeling of AHSR by conducting experiments with two most well-known algorithms for sequential data, DTW and HMM. The purpose of this study is to design a hand signal recognition system to be used in combat zone for more efficient and safer maneuver of small unit. Therefore, the commercialization of AHSR can be considered only if a certain level of accuracy and speed are satisfied; and the criteria should be relatively stricter than the other systems with intention to use in daily life. Therefore, the minimum criterion for accuracy was set at the level of 90%, and the experimental results for cases that do not meet the minimum criterion are not discussed in detail. In order to find optimal recognizer design, experiments were conducted with dataset from one military trained soldier; the data from the other participants were used to verify the design. All experimental results are determined by the mean values of the results from 50 times of trial conducted under the identical conditions. However, as the training and test data set were dynamically configured, the composition of each data extracted from training and test data is irrelevant.

The conditions that designer can choose for AHSR modeling are listed in Table 4. The following subsections will explain more details of each selectable and tunable condition and the experimental results based on that. According to each analysis, a section 4.3.6 will provide conclusions for optimal modeling.

Table 4. Selectable or tunable conditions for optimal modeling

Algorithm	**Feature**	**Number of training data**	**Model**	**Modeling Variable**
DTW	Accelerometer Gyroscope Linear Accelerometer Pre-processed data Combinations of above	5~55	-	-
HMM			Left-to-Right Ergodic	Sigma (σ) Down-sampling Committee size

4.3.1 Algorithm Selection

The DTW (Dynamic Time Warping) and HHM (Hidden Markov Model) are for core techniques of this study. Although they look similar theoretically, they have much different operations and results. This section will compare the DTW and HMM, and discuss different results varied by each adjustable setting. To focus on different performance of two algorithms in depth, other variables were fixed or broadly considered; the number of training data was fixed as 616 (44 per gesture), the maximum number of input data for this study. All results are the mean value calculated by repeating each experiment 50 times; i.e. the recognition accuracy is the mean value of 7700 times of test. At the end, the optimal decision in terms of algorithm will be suggested.

To start with conclusion, DTW didn't work well with 14 types of hand signal gestures; the recognition precision for 14 types of gesture was 53.13% in average. A potential reason is that the gestures are very similar to each other. DTW uses the training data to generate a template and the template is compared with input data for the classification. If each gesture for hand signals is divided into smaller pieces, different gesture can share the same pieces. Due to the operation principle of DTW and the similarity of gestures, the time warping can transform two different gestures into mostly identical templates. In order to find intersectionality of AHSR system and DTW's usability, the prediction accuracy was evaluated according to the number of gesture types. The number of gesture types was increased to 3, 5, 7, 10, 12, and 14, based on frequency of use. Each set of gesture types are summarized in Table 5.

Table 5. The set of gesture for the different number of gesture types; The bold numbers indicate newly appeared gesture ID from the previous row.

The number of gesture types	Gesture ID
3	**1, 2, 3**
5	1, 2, 3, **4, 10**
7	1, 2, 3, 4, **5, 6,** 10
10	1, 2, 3, 4, 5, 6, **7, 8,** 10, **12**
12	1, 2, 3, 4, 5, 6, 7, 8, **9,** 10, 12, **13**
14	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, **11,** 12, 13, **14**

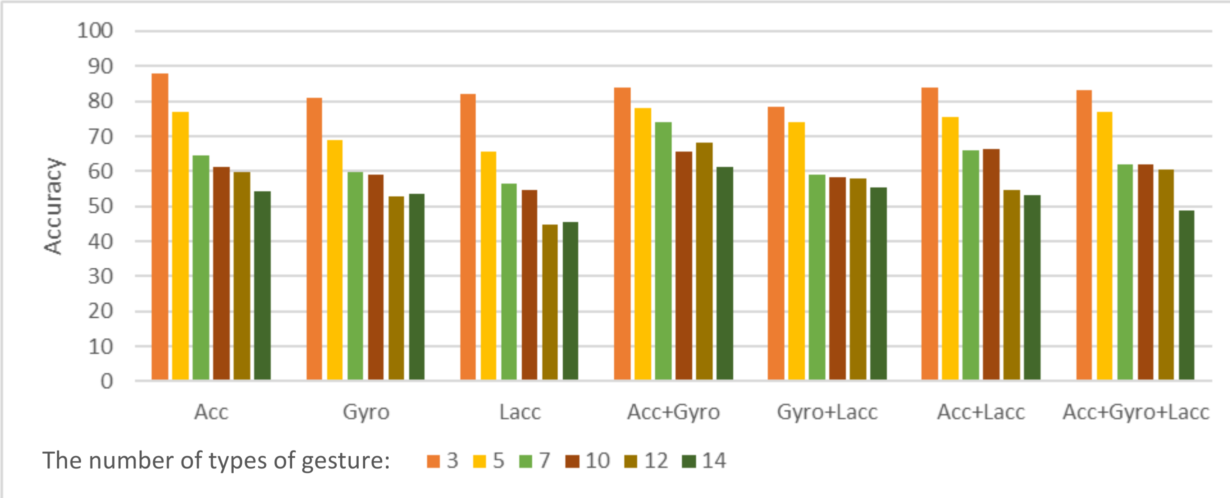


Figure 15. The mean value of prediction accuracy with increasing number of the gesture types

Figure 15 shows the mean value of prediction accuracy with increasing number of the gesture types for each feature and feature set[8]. The highest value of precision was 88.0% for 3 types of gesture with acceleration data, and the accuracy declined as the number of gestures increased. For the total set of 14 gesture, the maximum accuracy was only 61.039%, well below the minimum standard.

However, it is noteworthy that the prediction of DTW time was much shorter than that of HMM (about 1/3 of HMM), although the training time of DTW was longer than the HMM was. The short prediction time is a great advantage under a sufficiently convincing assumption which states that the training for recognizer is completed before the performance of an operation and only recognition is used during the operation. The insufficient accuracy is the result of using raw data, and if another feature that specifically represent each hand signals can be defined, the use of DTW may be more beneficial. However, this study focuses on the use of raw sensor data, so more study for the availability of DTW for AHSR will be left for future the study.

[8] From this section of this paper, acronym *'Acc'* will indicate accelerometer, *'Gyro'* for gyroscope, and *'LAcc'* for linear accelerometer for the feature types.

For the HMM, there are a set of tunable parameters that make the system to achieve better performance; a type of model (left-to-right or ergodic), sigma $(\sigma > 0)$, committee size $(> 0 \text{ and } \leq \max number\ of\ class)$, down-sample factor (> 0). There is no golden-standard for these parameters, because the impact of each parameter on the recognizer depends significantly on the type of input data. Even if the type of input data is narrowed from the 'pattern' to 'gesture', there is still no golden-standard, because the outcome still can be varied by the types of gestures. Therefore, to find optimal value set of adjustable parameters, a large amount of experiment should be preceded.

There is two mostly used models for HMM, the left-to-right and ergodic. Figure 16 shows the comparison between two models for each feature, when $\delta = 1, \sigma = 1$ and not using down sampling. The left-to-right model showed 5.847% lower prediction accuracy than ergodic model in average; the maximum difference was 9.1039% for acceleration data and minimum difference was 3.8961% for linear acceleration data. Therefore, the ergodic model is considered to be more powerful than the left-to-right for all types of sensor data for this study.

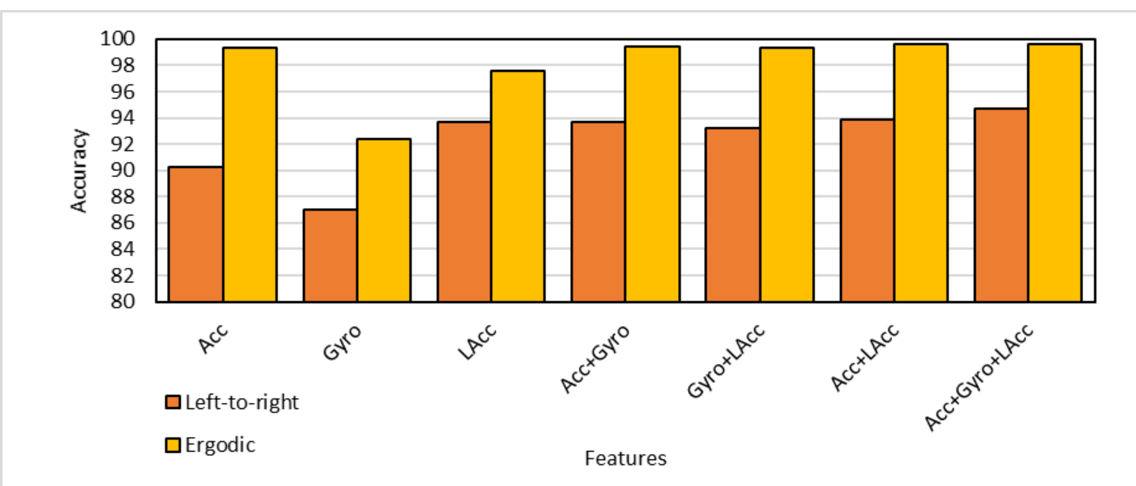


Figure 16. Mean recognition accuracy for left-to-right and ergodic model

The next tunable setting is sigma (σ) controlling how close each input vector needs to be to the model to be considered as a good match with the model. The σ was also decided empirically; with ergodic model and without down-sampling. As Figure 17 illustrates, the changes in precision according to the value of σ are different depending on the input feature; especially, gyro was particularly affected by σ. The acceleration and the angular speed have different range of value, as the movement of human arms, hands, and wrists present distinct bounds. Therefore, the σ should be separately considered for each feature, because the σ is affected by the range of dispersed value. As a result, σ shall be determined according to a set of features to be included in the final model for the recognizer.

The down-sample factor is used to get the gain in time cost, but always has the risk of data loss. The experimental results show that the accuracy was almost similar until the down-sample factor had a value of 5, but the accuracy was sharply reduced when it became larger than

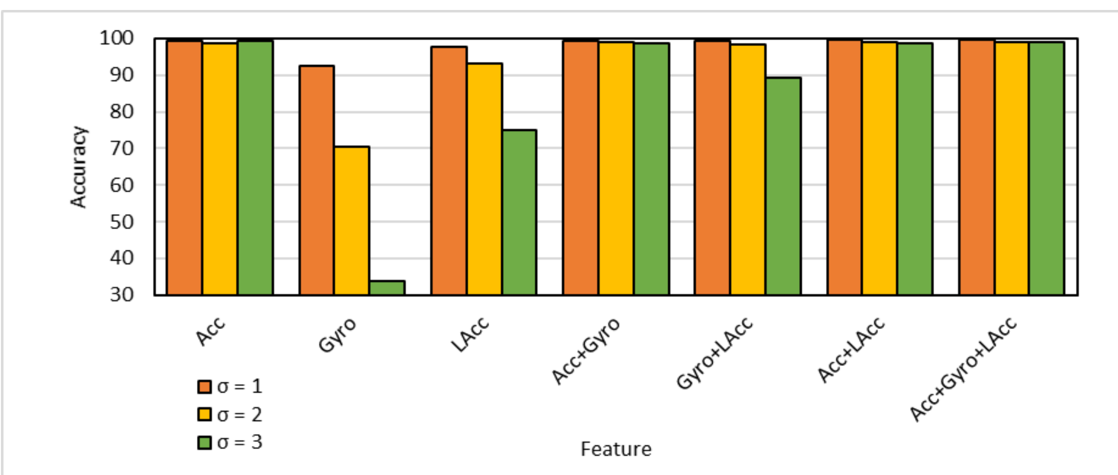


Figure 17. Mean recognition accuracy according to value of σ

5. For this study, the benefit of time cost resulting from down-sampling was minor and insignificant. Using the the committee size to control the number of models for prediction also did not seem noteworthy for this study.

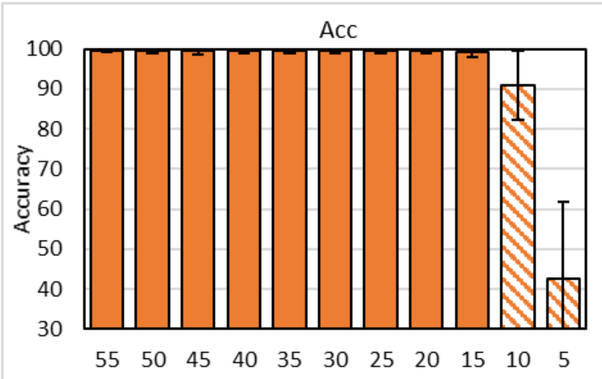
At this point, all features appeared to exhibit similar accuracy except Gyro. However, all features will be evaluated comprehensively with other elements using HMM in the following sections.

4.3.2 Number of Training Data

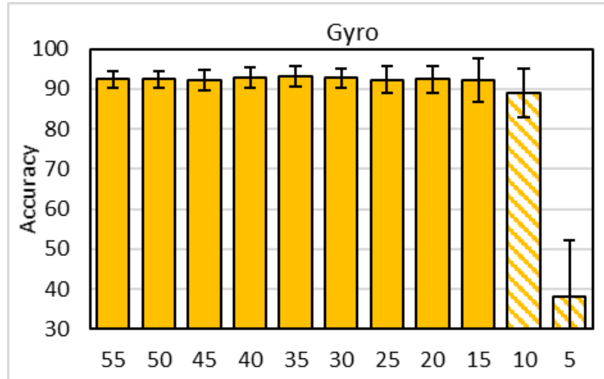
It is important to find the optimal size of required training data. Too small dataset can make the model difficult to achieve desirable accuracy, and too big dataset make hard for a user to use the system. Thus, deriving optimal size of data through experimentation is essential for both the system and the user. To evaluate only the effect of input data size on modeling, variables that were considered as irrelevant were fixed; the architecture of the recognition techniques, the value of σ, and the down-sample factor.

The experiments started with the maximum number of training data, 55 for each gesture. To make smaller set, the sample data was randomly selected using random API supported by Java and each set has reduced by 5 for each experiment.

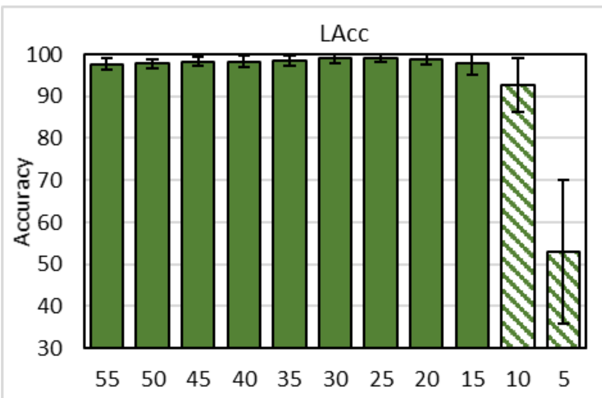
Figure 18 shows recognition precision according to the number of sample data for each set of features with HMM. The error bars indicate standard deviation of each case and the patterned bars mean the number of data is not enough for reasonable results. As mentioned before, the minimum criterion of accuracy was set as 90%, however, the mean value is not enough to make an optimal decision for the quantity of input data. Although the mean precision is greater than 90%, if the standard deviation is too big, the case should be examined considerably. For instance, for linear acceleration data with 10 training data (Figure 18c), the mean value of precision was 92.64% but standard deviation was 6.35. It means that the precision can be decreased significantly in worst case; the lowest precision among the 50 times of experiment with 10 LAcc data was 78.57%. With this respect, the system cannot be regarded as stable or robust system. All types of feature and feature set exhibited relatively sharp declining accuracy when the quantity of training data declined from 15 to 10. When the quantity of data was greater than 15, the precisions were not necessarily in direct proportion, and slightly fluctuated due to the quality of data set. Figure 19 shows the results of further exploration to find the most optimal number of data between 10 and 15. Repeating 10 to 15 times of gestures for the users is considered as affordable enough. However, it should be noted that the smaller number of data means the greater impact of training data’s quality.



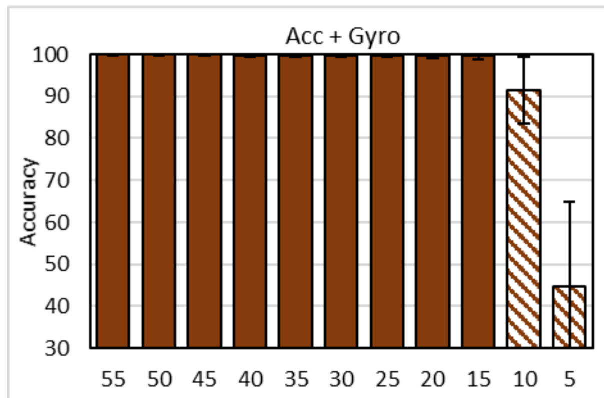
(a)



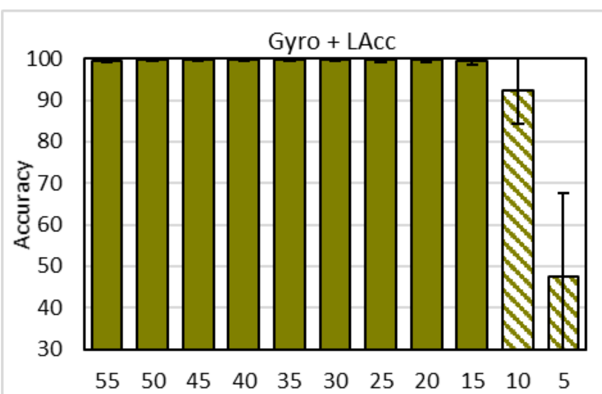
(b)



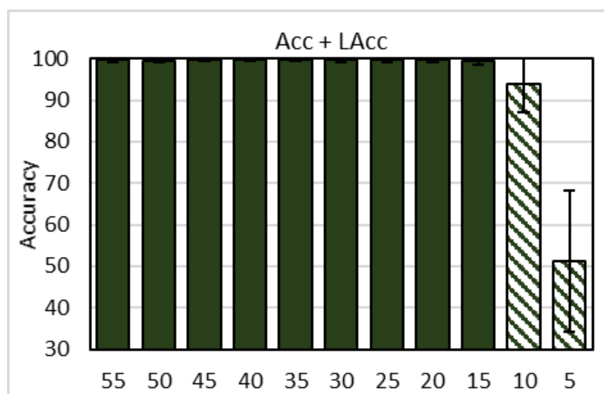
(c)



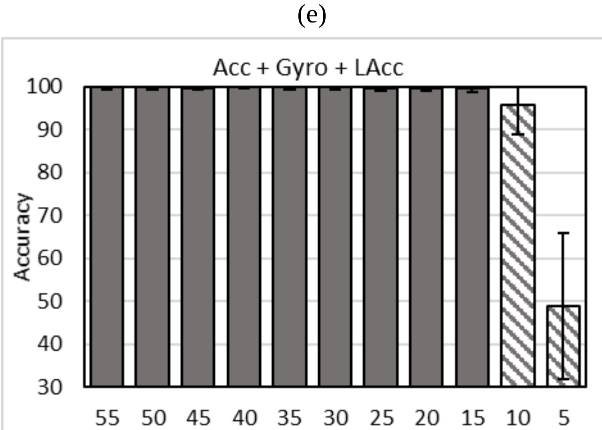
(d)



(e)



(f)



(g)

Figure 18. Mean recognition accuracy according to the number of training data with HMM for (a) Acc; (b) Gyro; (c) LAcc; (d) Acc+Gyro; (e) Gyro+LAcc; (f) Acc+LAcc; (g) Acc+Gyro+LAcc. The patterned bars indicate the number of sample data is not enough and the error bars represent the standard deviation of each results.

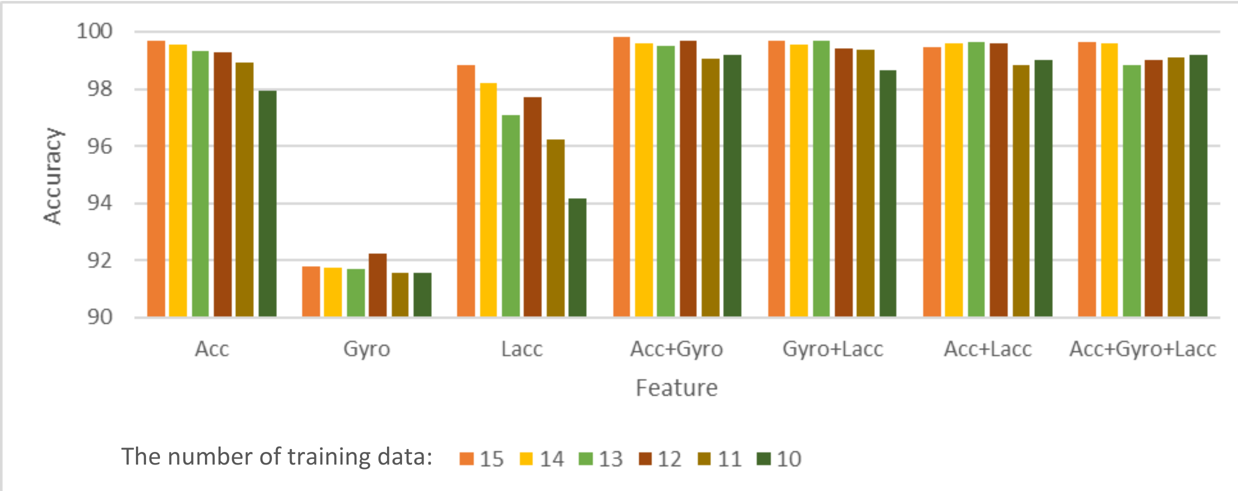


Figure 19. The mean value of accuracy according to the number of training data (10 to 15) for each feature

In Figure 19, the number of training data is further divided by the unit of 1, from 10 to 15. The Figure 19 provides some clear conclusions and more issues to be discussed. For example, in Figure 18 and 19, it firmly displays that the data from gyroscope only is not enough to represent each gesture. On the other hand, it should be discussed more in depth to decide the optimum combination of the number of training data and feature.

4.3.3 Feature Selection

Sensors that are frequently used in applications of gesture recognition are accelerometer and gyroscope. This study also concentrated on accelerometer and gyroscope, and the values of linear acceleration were additionally used to evaluate the effect of gravity on classification accuracy. At the early stage of this study, Fast Fourier Transform (FFT) of raw sensor data was also considered as a method for preprocessing. However, through reasonable amount of experiments, it appeared that the raw data is sufficient to represent each gesture and any other merits were not identified, but increase in time cost. Thus, three data types with three axes for each were recorded and evaluated.

The Figure 19 in previous section displays a compared information of different feature and feature set, as well as the different number of training data. First, the independent data of angular speed and linear acceleration are not sufficient to classify 14 kinds of hand signals, because precision values are significantly lower than those of other features. Acceleration data itself is also not an optimal decision, because it shows a relatively sharp fall when the number of training data decrease from 15 to 10. Furthermore, assuming that the AHSR will be used to recognize and classify more types of gestures later, the angular speed should be included for the system' robustness. With these regards, candidates for the optimal feature are the right 4 set from Figure 19, and the combination with the minimum number of the sample data are summarized in Table 6. Next sections will discuss with these four combinations.

Table 6. The candidate set of feature and the number of sample data for optimal design

Feature	Number of training data
Acc + Gyro	12
Gyro + LAcc	13
Acc + LAcc	12
Acc + Gyro + LAcc	14

4.3.4 Time Cost

For the AHSR system, operation speed should be considered very significantly; the time cost can be categorized into two, training time and prediction time. When considering a sufficiently convincing assumption which states that the training for recognizer is completed before the performance of an operation and only recognition is used during the operation, the prediction time is a obviously more significant.

The time cost evaluation was conducted with four types of feature set determined from the previous section in this paper, and Figure 20 describes the training and the prediction time for each in milliseconds. Because each feature possesses 3 axes data, 2 types of feature require a 6-dimensional recognizer and 3 types of feature require a 9-dimensional recognizer. Training and prediction with larger dimension are inevitably accompanied by longer operating time. As discussed in the feature selection part, as the accuracy are similar, using two types of feature is more advantageous in terms of the time cost than using three types of feature, even it is a very small difference (0.43 sec for training and 0.1 sec for prediction in average). In this situation, the linear acceleration data obtains one additional benefit. For the real-time recognition part of AHSR that will be introduced in the section 4.4 of this paper, the linear acceleration data can be intuitively used for gesture detection.

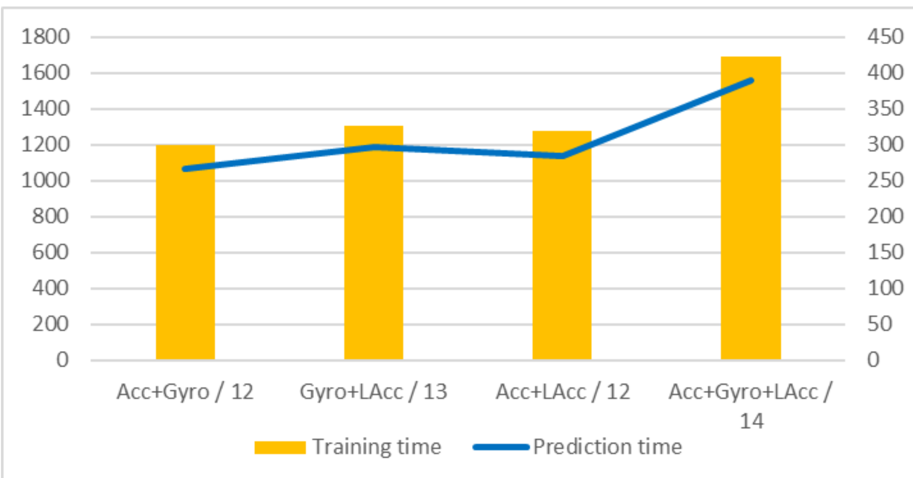


Figure 20. Training and Prediction time for each combination of feature set and the number of sample data in milliseconds. The bar chart with left vertical axis indicates the training time and the line chart with right vertical axis describes the prediction time.

In conclusion, the optimal design for AHSR modeling refers to:

- HMM with $\sigma = 1$
- Angular speed (Gyro) and linear acceleration (LAcc) data
- 13 training data for each gesture

This conclusion will be evaluated in the next section with data from 10 participants.

4.3.5 Evaluation

The objective of this section is to validate an optimal recognizer design which is decided through previous sections with 10 participants' data. Table 7 shows the results of accuracy, training time and prediction time for the designed recognizer. The value of accuracy and time cost was also calculated by 50 times repetitive process.

Table 7 Evaluation results for an optimal recognizer model with data from 10 participants

#	1	2	3	4	5	6	7	8	9	10	Avg.
Accuracy (%)	98.82	97.60	98.60	99.72	98.28	99.67	99.13	100.0	100	99.57	99.14
Training time (sec)	2.81	2.85	1.89	2.01	2.05	2.89	3.48	2.04	2.96	2.54	2.55
Prediction time (sec)	0.37	0.86	0.75	0.65	0.62	0.59	0.46	0.77	0.80	0.56	0.64

As Table 7 describes, the result of accuracy was 99.14% in average; maximum accuracy was 100.00% and minimum accuracy was 97.60%. This result seems to be high enough but somewhat lower than the results from the modeling step (99.69%). The main reason for the decrease in accuracy is that users were very unfamiliar with the movements for army hand signal. For most of participants (even if they answered that they had military experience in the questionnaire), it was their first time to perform the hand signals in their life. The participants did a little different gesture each time. They are also unware of the sensitivity of the smartwatch sensors, the detailed classification principles of the pattern recognition using HMM. Therefore, it is expected that higher accuracy will be obtained if a user is military-trained person and more tutoring about the system can be done in advance. Further experiments and evaluations with military-trained participants will increase validity of the statement.

4.3.6 Conclusions

The system showed an accuracy of 99.14% in average with mostly non-military trained people with limited knowledge in gesture recognition. With this restraint, accuracy of 99.14% and the time cost that does not deviate much from the expectation implies that the system is designed to be used for tactical purpose. In conclusion, this optimal design for recognizer can be applied to real-time AHSR.

4.4 ARMY HAND SIGNAL RECOGNITION SYSTEM

4.4.1 System Overview

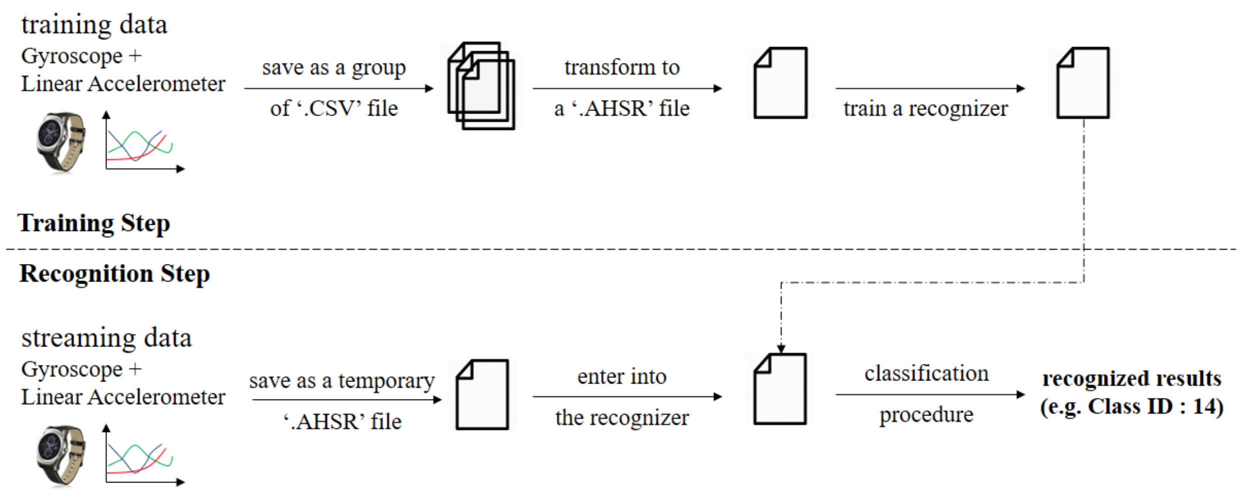


Figure 21. A conceptual diagram for real-time AHSR

Section 4.4 will introduce a real-time AHSR system. For the original purpose of this study, performance in a real-time environment is important. Figure 21 is a conceptual diagram including both steps of the training and the recognition of AHSR. The optimal design for the recognizer including algorithm and feature selection was discussed in previous section. To utilize these outcomes for the real-time system, some additional factors should be discussed; recording and managing of data in real-time, and extraction of proper gesture fragments from successive streaming data.

When the system detect a meaningful movement, it extracts data of a certain length of time and stores it in a temporary '.ahsr' file. The system classifies the gesture using the data file and reports the classification result (the predicted gesture ID). When the prediction is done, the file would be deleted for efficiency of a device's memory space. The gesture detection technique will be discussed in the following section.

4.4.2 Gesture Detection

It should be designed carefully for real-time gesture recognition that how the system determine which parts of the streaming data are to be extracted, analyzed and classified. A sophisticated algorithm for motion detection is not difficult to design and several respectable previous researches exist already. However, AHSR needs to handle short sequential data, which is different from a activity recognition, and it requires shorter operation time than normal, for the purpose of its usage. With this respect, the gesture detection and the extraction algorithm for AHSR should be fairly simple and take a relatively short time.

Therefore, raw linear acceleration data was used for gesture detection for this study. As described in section 4.1.2, linear acceleration data is not including gravity. In other words, when a device is not moving, all three axes return zero values. However, because the sensor is sensitive enough to catch the vibration of a person's hand that is not moving, it is very rare for all axes to exhibit zero values. So, the sum of squares of the values of the three axes can be used for gesture detection; in this part of paper the sum value would be referred as an 'energy'.

Furthermore, it is necessary to set two kinds of threshold for discriminating a meaningful gesture and insignificant small movements; an energy threshold and a time length threshold. First, an energy threshold is a criterion for determining how much the value of energy should be judged as meaningful motion. If the value of energy threshold is too small, the recognizer will be triggered too often or will not find the correct starting point of the gesture. On the other hand, if the value is too big, the system may not recognize a slow-running gesture, or may miss early data of a gesture. The value of threshold for energy determined to be 1.1 empirically.

Second, the time length threshold is a standard to make a system to recognize a movement as a gesture only if the energy lasts for a certain time with exceeding certain value.

The average, minimum, and maximum time length of all types of gestures were obtained through analysis of accumulated data. The time length threshold was fixed as a constant by using these statistical values. Fixing the time threshold to a constant value can simplify the gesture detection of the system, but can have negative impact on the robustness of the system if the user's movement speed largely varies each time. For this study, because the expected users of the system are trained combatants, the fixed value of time length threshold was judged to have more positive effects on the system and it was determined to be 15 data (meaning 0.3 seconds in average). The gesture detection procedure is summarized in Algorithm 1.

Algorithm 1. Gesture Detection Algorithm for AHSR
0: initialize all thresholds
1: read streaming data of Gyro (x, y, z) and LAcc (x, y, z) from smartwatch
2: energy = $LAcc\_x^2 + LAcc\_y^2 + LAcc\_z^2$
3: if energy is > 1.1
4: if the time length threshold is >15: start to record data on a file
5: if the time length of recorded data is > 3.3 sec: trigger the recognizer
6: initialize all thresholds

4.4.3 Evaluation

First of all, one more threshold had to be explored for real-time AHSR, an irrelevant gesture classification threshold. It is a threshold for preventing the system from classifying irrelevant hand signals as a significant movement. The value of likelihood for each class can be used for the irrelevant gesture classification threshold. If the threshold is set too high, there is a risk of not classifying valid hand signals; if set too low, meaningless movement can be classified as a signal. When the recognizer using HMM receives the data to analyze, it calculates likelihoods for each class, and the class number with the largest likelihood value is returned as the classification result. At this time, if the likelihood value of each class is similar and the maximum value is not high enough, the input movement should not be classified as a meaningful gesture. Therefore, the irrelevant gesture classification threshold is the standard for this situation.

Analyzing the results from optimal modeling step, when the recognizer correctly classified the input gesture, the maximum likelihood was distributed between 0.7 to 1.0 in average. Although the system rarely returned correct result with a likelihood less than 0.7, it was determined that it would make the system more sophisticated by discarding that kind of data and making the user to perform the same signal once again. The irrelevant gesture classification threshold was empirically decided as 0.68. In other words, when the system analyzes the input data, if the maximum likelihood value is less than 0.68, it discards the input data and do not return any result (for this study, it returned class ID 0 to check the result).

To evaluate the real-time system, each participant was requested to conduct 14 types of gesture successively in random order and during this continuous gesturing, data was streamed continuously. Each participant performed a set of 14 gestures only once. For this evaluation, it is difficult to evaluate iteratively to obtain the average result of accuracy. Instead, Table 8 summarizes the result. For 7 participants, the recognizer classified all 14 types of gesture correctly, but for 3 participants, there were 1 or 2 miss-classified gestures. In the cases that the system misclassified, two gestures were similar either the gesture’s characteristic was not distinct enough

Table 8. Experiment results for real-time AHSR with 10 participants

Participant # / Gesture #	1	2	3	4	5	6	7	8	9	10
1	√	√	√	√	√	√	√	√	√	√
2	√	√	√	√	√	√	√	√	√	√
3	√	√	√	√	√	√	√	√	√	√
4	√	√	√	√	√	√	√	9	√	√
5	√	√	√	√	√	√	√	√	√	√
6	√	√	√	√	√	√	√	√	√	√
7	√	√	√	√	√	√	√	√	√	√
8	√	√	√	√	√	√	√	√	√	√
9	√	4	√	√	√	√	√	√	√	√
10	√	√	√	√	√	√	√	√	√	√
11	√	√	√	√	√	√	√	√	√	√
12	√	√	√	√	√	√	√	√	√	√
13	√	√	√	√	√	√	√	√	√	√
14	√	√	√	√	-	√	√	-	√	√
Correct/False	14/0	13/1	14/0	14/0	13/1	14/0	14/0	12/2	14/0	14/0

4.4.4 Conclusions

The operational process of AHSR should be simple because of many environmental restraints, while the system requires a high degree of accuracy. Through repetitive experiments, the settings were adjusted, and the real-time AHSR with fast enough operational speed and high accuracy was designed. However, if more repetitive experiments with a larger number of participants can be done, more precise assessments of system performance will be possible.

5.0 DISCUSSION

5.1 LIMITATIONS

This paper is an early stage study demonstrating the feasibility of use of gesture recognition for military purpose, so there are several performance and study limitations that must be discussed. The limitations are introduced in this section, and the prospective solutions will be suggested in the next section.

First, this study deals only with the movement of one hand. However, for unrestricted expansion of the study, the classification and recognition of the movement of the other hand is suggested.

Second, proper compensation method for walking speed was not considered. It is legitimate to say that, in most case, a commander performs signaling in a stationary state for better visualization. However, the system will become more robust if hand signals performed during moving can take into.

Third, there was not enough evaluation with military trained participants. The statement that trained people will show higher accuracy with AHSR is still a hypothesis until more experiments are actually performed by them and the results to be come out. Furthermore, if this study can be obtained after the combatants have used the hand signals in actual reconnaissance operations, it will be great for the improvement of the system.

Currently, the AHSR requires data transmission from smartwatch to PC using WIFI. However, taking into a consideration on using it in the battlefield, the smartwatch to smartphone data transfer using Bluetooth can be more appropriate. Or, if all recognition and classification can be done on the smartwatch itself, it would be the best scenario.

Lastly, more efforts can be made to make the AHSR more user-friendly. The current version of the system is executed by using command lines, so it is difficult for non-expert to use it alone. If this study leads to development for practical use, efforts for the user interface must be accompanied.

5.2 FUTURE WORK

In this part, several aspects to complement limitations mentioned above and other predictable future work are introduced. Obviously, if further experiment and evaluation can be executed with military trained soldiers, that would be helpful to improve the performance of AHSR for the purpose of military use.

First, a small IMU sensors can be attached to non-dominant hand to recognize and classify the movements of both hands. Then the system can be used for more complex hand signals with both hands. However, as the dimension of the model to be analyzed increases, the time cost increases together. Therefore, efforts to create elaborate design for the system is required.

In terms of functionality, there are a few areas of improvements. An algorithm to compensate for walking speed can make the system more sophisticated. Also, there can be a specific gesture as a delimiter to cancel an already performed signal in order to prepare for an immediate situation change or to prepare for a combatant's mistake. For example, [25] suggested gesture delimiter for wrist-worn devices which is easy to perform and uncommon in most people's everyday lift; double Rotate. This kind of gesture which is unique and distinct from the hand signals can be used for AHSR. After sufficient functional improvement of AHSR, more efforts can be made to make the system more user-friendly.

There are many benefits if all of the functions of AHSR can be implemented in a smartwatch's stand-alone app. First, the data transfer is not required. Then time cost will decrease and the system can be more environment independent. If this is difficult due to some limited capabilities of smartwatch, communication between smartwatch and smartphone using Bluetooth can be an alternative way the successful implementation.

There was a cross-participants experiment to determine if a system can be provided with a rich gesture library that allows each user to skip the training procedure of a recognizer; in other words, a gesture library that can be used by anyone and the reasonable accuracy can be achieved. However, the value of precision was very different according to the combination of training data and test data; accuracy was wide ranging from 35% to 90%. Therefore, more study is needed to find user independent gesture library for AHSR.

The summary of the limitations and future work displayed in this paper is that there are still several questions to be asked and answered, but also prospective solutions. The derivation of these future work is another achievement of this study.

6.0 CONCLUSIONS

In this paper, a gesture recognition system, Army Hand Signal Recognition (AHSR) system, to be used for reading and classifying a combatant's hand signal was introduced. Through numerous repetitive experiments, the elaborate recognizer was designed. The optimal recognizer design has been validated with the experimental data from 10 participants. By showing high accuracy and fast operating speed, the AHSR presented a successful early stage for military purpose applications of gesture recognition.

Specifically, the AHSR maximized the benefits of the system in restricted battlefield situations by using smartwatch. And the AHSR classified 14 types of complex and compound hand movements rapidly and accurately. Although, for real-time use in field of battle, the system had to be somewhat simplified, this study shows that the power of the IMU sensors of the smartwatch and the HMM can be sufficient to overcome such limitations. It means that there are many possible applications in the future. This study is expected to be the foundation point of such applications.

APPENDIX A

HAND SIGNALS FOR MANEUVER OF REPUBLIC OF KOREA ARMY

Below table shows detailed description for ROKA's each hand signal for maneuver operation. Each command meant by each gesture was eliminated due to security purpose and distinct ID number was imposed instead.

Category	**ID**	**Description**
Posture	1	Holding a fist, raise one arm to the shoulder level, and bend it to make 'L' shape
	13	Sit down as low as possible while putting both hands on the helmet
	14	Put up one palm on the muzzle of a gun while holding the gun muzzle faxing up straight
Gesture	2	With a palm facing up, stretch one arm from back toward front drawing a semicircle
	3	Holding a fist, raise one arm up above shoulder height and go down to shoulder height repeatedly
	4	Stretch a palm on the side, facing a palm horizontally to the ground. Then lower it down to 45 degrees, raise up to 90 degrees repeatedly
	5	With the right palm facing forward, stretch one are above head height and draw a small horizontal circle overhead several times

	6	Bend one arm so that a palm faces toward the face of oneself and move a palm toward back and forth repeatedly
	7	Bend one arm so that a palm faces forward and move it back and forth repeatedly
	8	Holding the right fist, stretch right arm from the shoulder and pull repeatedly
	9	With the palm facing the ground, stretch one arm backward and move it up and down repeatedly
	10	Put a palm on the helmet and spread the arms horizontally
	11	Bend one arm and point fingertips downward and draw a small horizontal circle at the waist height repeatedly
	12	With a stretched hand, stretch an arm toward the direction one wants to leap

APPENDIX B

DEMOGRAPHIC SURVEY AND CONSENT FORM

The next page is the demographic survey and consent form that participants were asked to fill out.

University of Pittsburgh
School of Engineering
Department of Electronical and Computer Engineering

Weonji Choi
Master of Science Candidate
412-452-3301

Thank you for your participating.

This research is for my master thesis, tentatively titled **"Army Hand Signal Recognition System using Smartwatch Sensors"**. The organized armies of the world all have their own hand signal systems to deliver commands and messages between combatants during operations such as reconnaissance, and infiltration. For instance, to command a troop to stop, a commander would lift his/her fist next to the his/her face height. This study is aiming to develop a system that converts analog hand signal to digital signal, recognize and classify the signal to do further work. The smartwatch you've been asked to wear will send the data containing acceleration and angular velocity of your hand movement to the server.

You will be asked to fill up short questionnaire, and conduct 14 types of gesture repetitively. You can find a table describing each motion on the next page. After you finish this, please feel free to provide any comment to researcher for the system's sustainability. The whole experiment is expected to take about 0.5 - 1 hour. All information will be used only for research purpose and kept confidential.

Thank you for your time and efforts.

Please fill out following questionnaire.

Sex : Male / Female

Age : ___________

Height : ___________cm

What is your dominant hand? : Right / Left

Do you use a smartwatch? : Yes / No / In the past

Do you have military training experience? : Yes / No

- If yes, how long? _________ month / year

Do you have ever studied gesture recognition? : Yes / No

- If yes, please rate your experience from 1 (never) to 10 (very experienced): __________

By signing below, I agree to provide my information for this study.

(Date) ____________________________ (Signature) ____________________________

APPENDIX C

DEMOGRAPHIC INFORMATION OF PARTICIPANTS

Below table explains demographic information of each participant collected by using the form of appendix B.

#	Sex	Age	Height (cm)	Dominant Hand	Smartwatch User (Y/N)	Military Trained (Y/N)	Perceived awareness of Gesture recognition (0-10)
1	Female	25	164.0	Right	N	N	N
2	Male	38	170.0	Right	N	Y (2.2 years)	N
3	Male	30	175.1	Right	N	N	N
4	Female	30	172.0	Right	N	N	N
5	Male	30	175.0	Right	Y	Y (2.1 years)	N
6	Female	28	168.0	Right	Y	N	N
7	Female	25	167.2	Left	N	N	N
8	Female	35	163.1	Right	N	N	N

9	Male	38	176.0	Right	N	Y (2.2 years)	N
10	Female	27	160.5	Left	N	N	Y (4)

BIBLIOGRAPHY

[1] M. van Creveld, “Technology and War II, From Nuclear Stalemate to Terrorism,” in *The Oxford History of Modern War*, 2005, pp. 341–363.

[2] C. M. Bishop, *Pattern Recognition and Machine Learning*, vol. 4, no. 4. 2006.

[3] S. X. Ju, M. J. Black, S. Minneman, and D. Kimber, “Analysis of gesture and action in technical talks for video indexing,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 595–601, 1997.

[4] J. W. Davis and A. F. Bobick, “Virtual PAT: a virtual personal aerobics trainer,” *Work. Percept. User Interfaces - PUI ’98*, pp. 13–18, 1998.

[5] J. Crowley, F. Berard, and J. Coutaz, “Finger tracking as an input device for augmented reality,” *Int. Work. …*, no. June, pp. 1–8, 1995.

[6] G. A. Berry, V. I. Pavlovic, and T. S. Huang, “A Multimodal Human–Computer Interface for the Control of a Virtual Environment,” *AAAI Work. Intell. Environ.*, pp. 141–144, 1998.

[7] K. Katsuragawa, K. Pietroszek, J. R. Wallace, and E. Lank, “Watchpoint: Freehand pointing with a smartwatch in a ubiquitous display environment,” *Proc. Int. Work. Conf. Adv. Vis. Interfaces - AVI ’16*, pp. 128–135, 2016.

[8] S. Kratz, M. Rohs, and G. Essl, “Combining acceleration and gyroscope data for motion gesture recognition using classifiers with dimensionality constraints,” *Proc. 2013 Int. Conf. Intell. user interfaces - IUI ’13*, p. 173, 2013.

[9] M. Hoffman, P. Varcholik, and J. J. LaViola, “Breaking the status quo: Improving 3D gesture recognition with spatially convenient input devices,” *Proc. - IEEE Virtual Real.*, pp. 59–66, 2010.

[10] S. Yun, Y.-C. Chen, and L. Qiu, “Turning a Mobile Device into a Mouse in the Air,” *Proc. 13th Annu. Int. Conf. Mob. Syst. Appl. Serv. - MobiSys ’15*, pp. 15–29, 2015.

[11] P. Zhou, M. Li, and G. Shen, “Use It Free: Instantly Knowing Your Phone Attitude,” *Mobicom’14*, pp. 605–616, 2014.

[12] L. Kratz, M. Smith, and F. J. Lee, “Wiizards: 3D gesture recognition for game play input,” *Proc. 2007 Conf. Futur. Play Futur. Play 07*, pp. 209–212, 2007.

[13] Y. Dong, A. Hoover, J. Scisco, and E. Muth, “A new method for measuring meal intake in humans via automated wrist motion tracking,” *Appl. Psychophysiol. Biofeedback*, vol. 37, no. 3, pp. 205–215, 2012.

[14] H. Huang and S. Lin, “Toothbrushing Monitoring using Wrist Watch,” *Proc. 14th ACM Conf. Embed. Netw. Sens. Syst. CD-ROM - SenSys ’16*, pp. 202–215, 2016.

[15] F. Mokaya, R. Lucas, H. Young Noh, and P. Zhang, “MyoVibe: Vibration Based Wearable Muscle Activation Detection In High Mobility Exercises,” *Proc. 2015 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput.*, pp. 27–38, 2015.

[16] N. Friedman, J. B. Rowe, D. J. Reinkensmeyer, and M. Bachman, “The manumeter: a wearable device for monitoring daily use of the wrist and fingers,” *IEEE J. Biomed. Heal. informatics*, vol. 18, no. 6, pp. 1804–1812, 2014.

[17] S. Siddhpuria *et al.*, “Exploring At-Your-Side Gestural Interaction for Ubiquitous Environments To cite this version : HAL Id : hal-01654892 Exploring At-Your-Side Gestural Interaction for Ubiquitous Environments,” 2017.

[18] C. Xu, P. H. Pathak, and P. Mohapatra, “Finger-writing with Smartwatch: A Case for Finger and Hand,” *Proc. 16th Int. Work. Mob. Comput. Syst. Appl. - HotMobile ’15*, pp. 9–14, 2015.

[19] S. Kratz, M. Rohs, and D. T. Laboratories, “A $3 gesture recognizer: simple gesture recognition for devices equipped with 3D acceleration sensors,” *Proc. 15th Int. Conf. Intell. user interfaces*, pp. 341–344, 2010.

[20] L. Porzi, S. Messelodi, C. M. Modena, and E. Ricci, “A smart watch-based gesture recognition system for assisting people with visual impairments,” *Proc. 3rd ACM Int. Work. Interact. Multimed. Mob. portable devices - IMMPD ’13*, pp. 19–24, 2013.

[21] S. Shen, H. Wang, and R. Roy Choudhury, “I am a Smartwatch and I can Track my User’s Arm,” *Proc. 14th Annu. Int. Conf. Mob. Syst. Appl. Serv. - MobiSys ’16*, pp. 85–96, 2016.

[22] J. Liu, Z. Wang, L. Zhong, J. Wickramasuriya, and V. Vasudevan, “uWave: Accelerometer-based Personalized Gesture Recognition and Its Applications.”

[23] D. Ashbrook and T. Starner, “MAGIC: A Motion Gesture Design Tool,” *Chi ’10*, pp. 2159–2168, 2010.

[24] A. Parnami, R. Sadana, A. Gupta, Y. Li, G. Reyes, and G. Abowd, “Mogeste: Mobile tool for in-situ motion gesture design,” *UbiComp 2016 Adjun. - Proc. 2016 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput.*, pp. 345–348, 2016

[25] F. Kerber, P. Schardt, and M. Lochtefeld, "WristRotate-A Personalized Motion Gesture Delimiter for Wrist-Worn Devices," *Proc. 14th Int. Conf. Mob. Ubiquitous Multimed. - MUM '15*, no. Mum, pp. 218–222, 2015.

[26] T. Vilarinho *et al.*, "A combined smartphone and smartwatch fall detection system," *Proc. - 15th IEEE Int. Conf. Comput. Inf. Technol. CIT 2015, 14th IEEE Int. Conf. Ubiquitous Comput. Commun. IUCC 2015, 13th IEEE Int. Conf. Dependable, Auton. Se*, pp. 1443–1448, 2015.

[27] M. Shoaib, S. Bosch, H. Scholten, P. J. M. Havinga, and O. D. Incel, "Towards detection of bad habits by fusing smartphone and smartwatch sensors," *2015 IEEE Int. Conf. Pervasive Comput. Commun. Work. PerCom Work. 2015*, pp. 591–596, 2015.

[28] I.-S. Oh, *Pattern Recognition*, 1st ed. Kyobo, 2008.

[29] B.-H. Juang, "On the Hidden Markov Model and Dynamic Time Warping for Speech Recognition-A Unified View," *AT&T Bell Lab. Tech. J.*, vol. 63, no. 7, pp. 1213–1243, 1984.

[30] S. Nakagawa and H. Nakanishi, "Speaker-Independent English Consonant and Japanese Word Recognition by a Stochastic Dynamic Time Warping Method," *IETE J. Res.*, vol. 34, no. 1, pp. 87–95, Jan. 1988.

[31] J. C. Lucero, K. G. *Munhall, V. L. Gracco, and J. O. Ramsay, "On the Registration of Time and the Patterning of Speech Movements," J. Speech Lang. Hear. Res., vol. 40, no. 5, p. 1111, Oct. 1997.*

[32] *H. Li, W. Yang, J. Wang, Y. Xu, and L. Huang, "WiFinger : Talk to Your Smart Devices with Finger-grained Gesture," Proc. 2016 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput. - UbiComp '16, pp. 250–261, 2016.*

[33] *S. Theodoridis, A. Pikrakis, K. Koutroumbas, and D. Cavouras, Matlab Introduction to Pattern Recognition: A MATLAB Approach, 1st ed. ACADEMIC PRESS, 2009.*

[34] *S. Axelrod and B. Maison, "Combination of Hidden Markov Models with Dynamic Time Warping for Speech Recognition," Proc. 2004 IEEE Int. Conf. Acoust. Speech, Signal Process. (ICASSP '04), p. 1–I, 2004.*

[35] *L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," Proc. IEEE, vol. 77, no. 2, pp. 257–286, 1989.*

[36] M. Barry, J. Gutknecht, I. Kulka, P. Lukowicz, and T. Stricker, "From motion to emotion: a wearable system for the multimedial enhancement of a butoh dance performance," *J. Mob. Multimed.*, vol. 1, no. 2, pp. 112–132, 2005.

[37] X. Wang and G. Dai, "A novel method to recognize complex dynamic gesture by combining HMM and FNN models," *Proc. 2007 IEEE Symp. Comput. Intell. Image Signal Process. CIISP 2007*, no. Ciisp, pp. 13–18, 2007.

[38] Y. Nam and K. Wohn, “Recognition of Space-Time Hand-Gestures Using Hidden Markov Model,” *ACM Symp. Virtual Real. Softw. Technol.*, no. MAY 1997, pp. 51–58, 1996.

[39] R. I. Ramos-Garcia and A. W. Hoover, “A Study of Temporal Action Sequencing During Consumption of a Meal,” *Proc. Int. Conf. Bioinformatics, Comput. Biol. Biomed. Informatics - BCB’13*, pp. 68–75, 2007.

[40] N. Gillian and J. a Paradiso, “The Gesture Recognition Toolkit,” *J. Mach. Learn. Res.*, vol. 15, no. November, pp. 3483–3487, 2014.