Rhythmic-based Dynamic Hand Gesture Authentication using Electromyography (EMG)

Submitted to
Graduate School of Information Science and Technology
Osaka University

July 2022

Alex WONG Ming Hui

Publications

I. Journal Paper

1. Wong, A.M.H.; Furukawa, M.; Maeda, T. Robustness of Rhythmic-Based Dynamic Hand Gesture with Surface Electromyography (sEMG) for Authentication. *Electronics* 2020, *9*, 2143. https://doi.org/10.3390/electronics9122143. (**Dissertation Chapter 3**)

II. International Conference Paper

1. Wong, A.M.H.; Furukawa, M.; Ando, H.; Maeda, T. Dynamic Hand Gesture Authentication Using Electromyography (EMG). In Proceedings of the 2020 IEEE/SICE International Symposium on System Integration (SII), Honolulu, HI, USA, 12–15 January 2020; pp. 300–304. (**Dissertation Chapter 4**)

Abstract

Authentication is the process of proving an individual to be genuine. In conventional authentication (mostly static authentication), an identifier (username / identity) and a verifier (password / evidence) are required to authenticate an individual. But if both identifier and verifier are compromised, authentication can be achieved by impostors. In dynamic authentication, an authentication protocol that changes every now and then (one-time password / behavior) is used to validate the identifier and sometimes verifier. This can be very difficult for the impostor to steal or mimic as it is always different.

One type of dynamic authentication is known behavioral biometrics. However, it can be inconsistent as it does not stay constant all the time. Therefore, in my research, I have proposed rhythmic-based dynamic hand gesture. The rhythm of the gesture is categorized as the content or ownership factor depending on the use case, while the gesture behavior is the style. Rhythm is chosen due to the ease of memorization (in both short and long term) as compared to words or numbers, and also acts as the static part of the authentication which provides stability and consistency in the performance gesture as it has indicators of the starting, action, ending point of the gesture. This combination forms dynamic authentication.

Hand gesture is a more natural way in communication and has more variation in gesture movements. Hand gesture can be analyzed with different techniques and the most popular is using camera. But camera-based hand gesture has a number of disadvantages such as limited field of view, lighting, space, and others. Therefore, in my research, I have chosen electromyography (EMG) which is a technique in evaluating muscle activity through electrical signals. The advantages of using EMG are convenience (wearable and wireless), less obstacles (direct contact to the body), and can be performed without spatial concern (only needs to contract and relax muscles). Throughout my experiments, the device that I have used is a wearable EMG armband known as Myo Armband. This device consists of 8 channels of EMG sensors circulating the forearm to retrieve electrical signals from the movement of the muscles.

In my research, there are two use cases for rhythmic hand gesture, user-dependent rhythm and one-time rhythm. User-dependent rhythm is when the user creates their own rhythm, much like a password but dependent on the content and style of the rhythmic gesture. One-time rhythm is when rhythm is generated by a generator much like one-time password, making the rhythm an ownership factor, while the hand gesture that performed the generated rhythm is an inherence factor, making one-time rhythm a multi-factor authentication.

For rhythmic hand gesture to be viable as a biometrics authentication, it has to achieve the desirable biometrics characteristics. These characteristics are universality, uniqueness, permanence, collectability, performance, acceptability, and circumvention. Different real-world scenarios (sitting, standing, walking), time period, and mimicry have been used to test out the aforementioned characteristics. Yielding an equal error rate of

up to 1.32% through different scenarios and time period have proven that rhythmic hand gesture has permanence and performance; whereas after mimicking, the false acceptance rate (FAR) is as low as 10% which compare to the FAR of a compromised fingerprint system which is 67% [1] proves that rhythmic hand gesture has uniqueness and circumvention.

One-time rhythm has been proposed in this dissertation to increase security of rhythmic hand gesture. Reason being that generated rhythm is always different. One-time rhythm requires both the recognition of rhythm and the recognition of behavior from the rhythmic gesture in the verification process. The system will first assure that the rhythmic gesture performed matches the rhythm generated. If the rhythm matches, it then proceeds to recognize whether the rhythmic gesture behavior is that of the intended person. The combination of Levenshtein distances and cross-correlation has been used to match the rhythms; whereas sliding window methodology has been used to detect subject-dependent features, which also allows the use of different rhythms. Different rhythms can have similar parts which allows similarity in behavior during the gesture performance. The overall performance of one-time rhythm uses product-rule of score-level fusion which reduces the false acceptance rate. Although shadowing methodology has been used in one-time rhythm, it has been proven to require little to no learning curve.

Rhythmic hand gesture as biometrics has proven to be a robust biometric authentication method. This has been investigated through user-dependent rhythm in different scenarios and time period with equal error rate of 1.32%. One-time rhythm is able to increase security of rhythmic hand gesture through multi-factor through the independency of rhythm recognition and behavior recognition.

Table of Contents

Introduction 1

1.1 Desirable characteristics in biometric authentication 1

1.2 Convenience, reliability, and security 2

1.3 Dissertation outline 3

Rhythmic-based dynamic hand gesture authentication using electromyography 5

2.1 What is authentication? 5

2.2 Authentication factor 5

2.2.1 Main authentication factors 5

2.2.2 Multi-factor authentication 6

2.3 Biometrics 7

2.3.1 What is biometrics? 7

2.3.2 Physiological biometrics and behavioral biometrics 8

2.3.3 Multi-factor authentication with biometrics 9

2.4 What is rhythmic-based dynamic hand gesture? 10

2.4.1 Rhythm as static password 10

2.4.2 Electromyography (EMG) 12

Rhythmic-based dynamic hand gesture: a robust biometrics 16

3.1 Robustness of rhythmic-based dynamic hand gesture 16

3.1.1 Experimental setups 16

3.1.2 Data Processing 18

3.1.3 Experimental conditions 21

3.1.4 Biometric characteristics of rhythmic-based hand gesture 27

3.2 Comparison with other biometrics 29

3.3 Summary 31

One-time rhythm 36

4.1 What is one-time rhythm? 36

4.1.1 External visual and auditory cues 37

4.1.2 Experimental setups 39

4.2 Rhythm and Behavior 40

4.2.1 Rhythm recognition 40

4.2.2 Behavior recognition 47

4.2.3 Rhythm Recognition × Behavior Recognition 57

4.2.4 No learning curve required for users 58

4.3 User-dependent rhythm versus one-time rhythm 59

4.4 Summary 60

Conclusion 62

5.1 Limitation and future work 63

Chapter 1

Introduction

1.1 Desirable characteristics in biometric authentication

Since ancient times, different methods of authentication had been used to identify and verify a person from merely recognizing a person to communicate through secret words or owning a unique item. Despite being ancient, these methods are still being used daily through digital means. But one thing always remains throughout, bad actors trying to gain access as genuine individual through stealing or impersonation.

As authentication in the digital world improves, so does the technique used to impose. Biometrics has recently gained traction in digital authentication due to its conveniences and uniqueness. Similar to the aforementioned "recognizing a person", biometrics is the measurement and analysis of a person's physical and behavioral characteristics. Computer takes more detail measurements to recognize a person which a normal human cannot detect with just their senses alone.

Any human physiological or behavioral characteristics can be considered as biometrics on condition that they have the following desirable properties [2]:

i. Universality – availability for the majority of the people
ii. Uniqueness – no two persons have the same in terms of characteristics
iii. Permanence – resistance to changes
iv. Collectability – characteristics are easy to acquire and quantifiable
v. Performance – how well it functions
vi. Acceptability – whether people are willing to accept the biometric system
vii. Circumvention – how easy can the system be tricked by fraudulent techniques

The aforementioned characteristics were first coined for human identification in 1994 [3], which included other characteristics that have been omitted due to insignificancy or outdated, such as:

- Indispensability – since biometrics is part of a human, it is always a natural characteristic
- Storability – as technology advances, storage size and compression method has improved to the point where storability becomes insignificant
- Exclusivity – as fraudulent techniques improve, only one form of identification is increasingly difficult to maintain security; thus, multi-factor authentication (MFA) has been introduced which require more than one form of identification
- Precision – this is very similar to uniqueness and performance
- Cost – advancement in technology and demand have lower the cost for biometrics
- Convenience – advancement in technology has enabled the measuring and storing the biometrics to be insignificant in terms of time

Although it is mentioned that convenience characteristic from the point of system is insignificant, convenience in terms of human usage still plays an important role in biometrics, especially in conjunction with the acceptability characteristic. Henceforth, the term "convenience" or "convenient" in the dissertation will be pointing towards how convenience it is for people to use the biometric system.

All the aforementioned properties are only desirable meaning having them can improve both security and usability. But most biometric systems usually have certain compromises in terms the aforementioned properties. For example, fingerprint has low circumvention as it can be easily fooled by using replicated fingerprint [1], or facial recognition can sometimes be less unique especially for twin siblings [4], or gait has low permanence as it can be easily affected by terrain and clothes worn [5].

A novel biometric authentication – rhythmic-based dynamic hand gesture has been presented in this dissertation that optimizes most of the desirable properties, such as low equal error rate (EER) of 1.32% for *performance*, ability to be performed in different scenarios and over period of times which shows *permanence*, and others, including convenience characteristics. An in-depth analysis and discussion of rhythmic-based hand gesture will be evaluated in Chapter 3.

1.2 Convenience, reliability, and security

From the point of view of user, biometric authentication has to be:

i. Convenience – the biometric system should be easy to use
ii. Reliability – produce consistent result on different environmental and time factors
iii. Security – the biometric system should be able to tell genuine from impostors

These three features are not newly defined properties but instead a consolidation of the biometrics desirable properties mentioned in Chapter 1.1. Reliability consists of *permanence* and *performance*; whereas security consists of *uniqueness* and *circumvention*. Convenience on the other hand can be derived from *universality* and *collectability*. All these three properties can lead to *acceptability*.

Reliability can also be seen from the performance metrics of false rejection rate (FRR), which is the rate of genuine user being rejected by the biometric system. Higher FRR indicates that the biometric modality is low in *performance* and *permanence*. The opposite of FRR is known as true acceptance rate (TAR), which is the rate of genuine user being accepted by the biometric system. FRR can be affected by factors such as failure-to-enroll (FTE) where the user population for whom the biometric system fails to extract usable information from biometric sample, and failure-to-acquire (FTA) where

the verification or identification attempts has fail to capture sample with sufficient quality [6].

Security on the other hand can be derived from the performance metrics of false acceptance rate (FAR), which is the rate of impostors being falsely accepted by the biometric system as a genuine user. Higher FAR indicates that the biometric modality can be easily circumvent or low in security. The opposite of FAR is true rejection rate (TRR), which is the rate of impostors being rejected by the biometric system. Other factors that can affect security are how easy can it be stolen or mimic physically, and how unique is the biometric itself.

As for convenience, it is almost impossible to accurate measure how convenient a biometric system or modality is as it is subjective and dependent on the use case. Some biometric can be convenient to a person but may be inconvenient for another. For example, people with less profound fingerprint or dry skin may find fingerprint authentication to be inconvenience due to the higher rate of FTE and FTA. High FRR also affects the convenience of a biometric system as it can require the users to perform the biometric identification or verification multiple times.

However, it is almost impossible to improve all these three properties as there are tradeoffs between convenience, reliability, and security. The most well-known tradeoff in biometric authentication is the tradeoff between FAR and FRR. This tradeoff can be seen when trying to increase the chance of accepting genuine user, the system will also increase the chance of accepting impostors as genuine user; whereas to decrease the chance of accepting impostors, it can also decrease the chance of accepting genuine user. This is reflected in receiver operating characteristics (ROC) curve. Therefore, a balance between the FAR and FRR is required. One method to do so is to find equal error rate (EER) of the biometric modality. EER is the point where FAR equals to FRR. The lower the EER, the better the overall performance of the biometric modality.

These three properties will be discussed as part of the problem statement for rhythmic-based dynamic hand gesture.

1.3 Dissertation outline

Chapter 2 will first introduce authentication. This topic covers the explanation of authentication, authentication factors which includes multi-factor authentication (MFA), and biometrics. Biometrics will be categorized into physiological / static biometrics, behavioral / dynamic biometrics, and multi-factor authentication with biometrics which leads to the proposal of one-time rhythm in Chapter 4. The proposed biometrics, rhythmic-based dynamic hand gesture will be introduced in this chapter. Electromyography (EMG) will also be introduced and discussed how it will be used and

the reason of choosing it for rhythmic-based dynamic hand gesture as compare with other techniques.

After discussing rhythmic-based dynamic hand gesture using EMG in Chapter 2, Chapter 3 discusses whether the proposed biometric modality is a robust biometrics by testing it on different scenarios, period of time, and mimicry. The performance metrics used in the test are FAR, FRR, and EER which reflects how well the system performance and how much it achieves has achieved in the desirable properties. The last part of the chapter will be the comparison of rhythmic-based dynamic hand gesture using EMG with other behavioral biometrics in terms of their advantages, limitations, and performance.

Chapter 4 introduces one-time rhythm. This chapter will discuss what and how one-time rhythm works, the relationship between rhythm and behavior and how they are independent from each other during the verification process. This chapter also provide evidence that different rhythmic gesture is able to produce similar TAR as same rhythmic gesture in recognizing behavior. The performance of one-time rhythm uses product rule of score-level fusion which fuses the performance value of rhythm recognition and behavior recognition which will explain the tradeoff of FAR and TAR in one-time rhythm. The final part of this chapter compares user-dependent rhythm from Chapter 3 with one-time rhythm on their use cases, advantages and disadvantages.

Chapter 5 concludes the dissertation with the findings of the research, goals achieved, and the current limitations of the proposed rhythmic-based dynamic hand gesture using EMG.

Chapter 2

Rhythmic-based dynamic hand gesture authentication using electromyography

2.1 What is authentication?

Personal data is getting invaluable but at the same time vulnerable. As technology improves, people are relying more on the technology. In return, these technologies require users to input more personal data to help improves the quality of the technologies the users used. Some of the personal data uploaded can be sensitive data and require utmost security to prevent the data from stolen. To protect these data, authentication has been used to correctly identify and verify the user to be genuine before authorization is provided to access these data.

Authentication is the act of verifying an individual's identity, to prove that they are who they claim to be. There are many ways to authenticate an individual; it can be in analog form or digital form, from presenting a passport to providing a password. But as technology evolves, so do the methods of authentication.

Since there are bad actors trying to steal or mimic other individuals' identity to obtain their information, the security of authentication has to be constantly improved to prevent identity theft. Enhancement to the existing authentication methods or new authentication methods that improve the security have to be proposed; but at the same time, it should not hinder or complicate the process of how user authenticate themselves. For example, to login into an account, the system requires the user to visit the company in person, along their identification credentials. Surely the example given is secure but it is infeasible for a daily use case. Therefore, recent authentication methods have been revolving around digital authentication which is available in almost everywhere at any time.

2.2 Authentication factor

2.2.1 Main authentication factors

Authentication especially digital authentication has three major factors [7]–[11]:

- knowledge factor: something that the user knows (e.g., password, PIN)
- ownership factor: something that the user possesses (e.g., token, key)
- inherence factor: something that the user is or does (e.g., fingerprint, gesture)

Knowledge factor is currently the most commonly used form of authentication. It requires the user to provide a secret of what the user knows in order to authenticate. This can be words or phrases in the form of speech, writing or typing. In computing, secret is usually input in the form of characters, digits, and symbols. A password is a secret string of characters used in authentication. As of writing, password is the most commonly used mechanism of authentication due to its simplicity and availability. Another popular form of password is the personal identification number (PIN), which is purely numeric. PIN consist of short string of numbers (usually between 4-8 digits) and is commonly used in monetary transaction. Traditionally, passwords are expected to be memorized; but recently there are applications that help managing passwords.

Ownership factor is a form of authentication using a key to a lock mechanism. The principle is that both the key comprises of a secret that is shared between the key and the lock. The key can be physical form and digital form. A passport is an example of a physical form; whereas digital signature is an example of a digital form.

Inherence factor usually uses biometric method to authenticate user. This can be the physical properties of the user, including fingerprint, face, iris, and others; or behavior properties of the user, such as gesture, gait, keystroke dynamics, etc. Biometrics will be discussed in more detail in the Chapter 2.3.

2.2.2 Multi-factor authentication

Authentication factors can be used in conjunction with each other. This is known as multi-factor authentication (MFA). MFA was proposed to increase safety and provide continuous protection from unauthorized access by using two or more credentials [11]. This means that a mixture of two or more factors are required to complete an authentication procedure. A common example of MFA is the combination of traditional password and one-time password (OTP) [9], [12]–[14]. This combination requires the user to input a password (knowledge factor) and an OTP that is generated from a device that the user possess (ownership factor).

MFA should provide more than two layers of security as the authentication methods used in the MFA are independent from each other during the verification process. When the first authentication factor in multi-factor authentication has been compromised, the second authentication factor should come in as an extra security to prevent the impostor from proceeding. In MFA, if either one of the authentication methods fails, the entire verification process fails.

Most MFA currently uses two-factor authentication (2FA) in the form of static-static authentication, i.e., both authentication methods stay the same or does not change within a short period of time, and static-dynamic authentication, i.e., one authentication method stays the same or does not change within a short period, while the other authentication method is not always the same or changes within a short period of time. Examples of static-static are passport and fingerprint, card and PIN; whereas examples of static-dynamic are password (static) and OTP (dynamic), card (static) and signature (dynamic).

2.3 Biometrics

2.3.1 What is biometrics?

The conventional authentication method in the digital world uses username combine with password or PIN. But as electronic devices are getting more compact and mobile, yet more powerful, biometric authentication has becoming more preferable as the de facto method for authentication in these devices.

Biometrics are measurements and calculations related to human characteristics and can come in many forms. Biometric authentication is the security processes of verifying an individual using their biometric traits. Biometric authentication is convenience as it is always available on the users since it is part of the users. There are two major biometrics – physiological biometrics and behavioral biometrics. Physiological biometrics, sometimes known as static biometrics, are based on the analysis of physical properties of a person. Examples of physiological biometrics are fingerprint, iris, facial recognition, and others. Behavioral biometrics, sometimes known as dynamic biometrics, are based on the analysis of behavior of a person in the process of reproducing an action. Examples of behavioral biometrics are gait, dynamic gesture, keystroke dynamics, and others.

Physiological biometrics are convenience and can be unique from person to person, but they are prone to presentation attack [15], [16]. This poses serious challenges as physiological biometrics do not change; thus, once physiological biometrics is stolen, the system will have difficulty to prevent the attackers from authenticating as the genuine user.

On the other hand, behavioral biometrics has an advantage over physiological biometrics in terms of security as behavioral biometrics decreases the chance of being stolen due to the innate behavior of every individual. Behavior can be different from one person to another, thus making it very difficult for impostors to perfectly mimic another person's behavior. Due to the difficulty in mimicking another person's behavior, behavioral biometrics have gained significant interest in biometric authentication. Behavioral biometrics can be inconsistent due to the ever-changing nature of behavior. Despite that, a person's behavior is not completely different when compared to previous behavior, unless done deliberately.

Biometric system can be evaluated through different performance metrics, such as confusion matrix, false acceptance rate (FAR), false rejection rate (FRR), area under the curve (AUC) of receiver operating characteristics (ROC) curve, and equal error rate (EER). The equations FAR and FRR [17] are as follows:

$$FAR = \frac{FP}{TN + FP}$$

$$FRR = \frac{FN}{TP + FN}$$

ROC curve is a graph that illustrates the performance of classifiers by presenting the trade-off between true acceptance rate (TAR) which can obtained from 1 – FRR, and FAR while varying the threshold. AUC of the ROC curve represents the degree or measure related to the classifier discrimination performance. It tells how much the model is capable at distinguishing between classes. Lastly, EER reflects the overall performance of a biometric system by predetermining the threshold value for FAR and FRR when they are equal. The closer the EER is to 0, the higher the performance of the system. EER can also obtain from the ROC curve as shown in Figure 2.1.

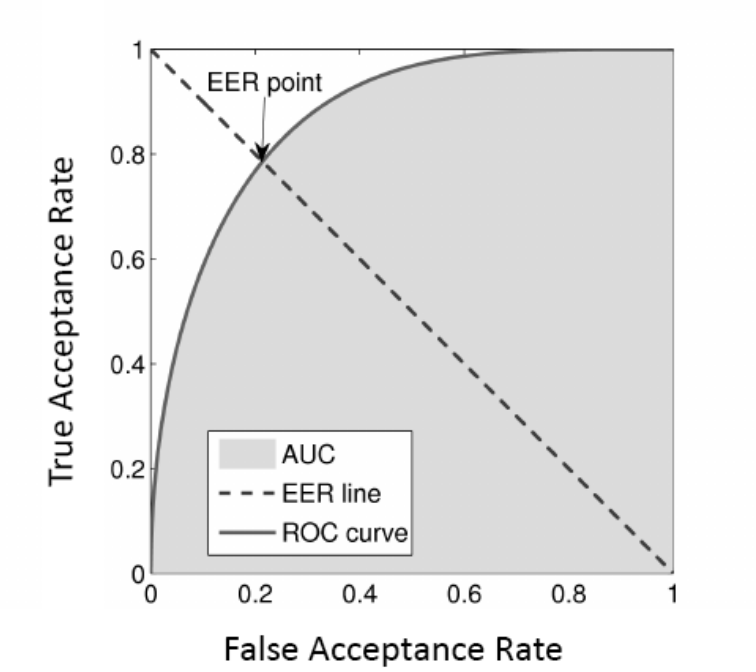


Figure 2.1: AUC and EER obtained from a ROC curve

2.3.2 Physiological biometrics and behavioral biometrics

Physiological biometrics, also known as static biometrics, is the measurement and analysis of what makes an individual the individual. Examples include but not limited to fingerprint, face recognition, iris recognition, body parts physical measurement, DNA, and even body odor. Physiological biometrics is the most commonly use biometrics due to its simplicity and convenience. These types of biometrics are unique and less likely to change within a short period of time as compared to behavioral biometrics, making them easy to use for authentication. But due to its permanence, it can be deemed to be less secure especially when the biometrics is compromised or stolen. For example, impostor is able to use an image of the target user to get pass through some facial recognition authenticator [18].

Behavioral biometrics, also known as dynamic biometrics, is the analysis and evaluation of pattern or behavior of how an individual performs certain action. This includes but no limited to typing rhythm as known as keystroke dynamics [19]–[21], gait,

dynamic hand gesture, signature, and voice. Most behavioral biometrics are known to inconsistent as it changes every time. But due to its ever-changing nature, it is also considered much difficult to be impersonate or mimic. One method to stabilize this inconsistency is by introduce a content to the behavioral biometrics. Example of these behavioral biometrics with content are dynamic signature verification, where the word sign is the content while the style of writing is the behavior.

Although, there have been research on cancellable biometrics [22]–[24] that can help improve the security or reduce the chance of impostors stealing the biometric data from the system level through biometric salting, e.g., encrypting biometric data, or non-invertible transformation, e.g., distorting stored biometric template, it still does not prevent the impostors from stealing the original biometric data physically or from the sensor level. Moreover, these cancellable biometrics can not only be integrated with physiological biometrics, it can also be implemented similarly to behavioral biometrics [25], [26], making them having the same advantages when it comes to cancellable biometrics. But in physical or sensor level attack, behavioral biometrics have the advantage due to its non-constant changes which can be difficult to steal, mimic, or impersonate.

2.3.3 Multi-factor authentication with biometrics

To increase conveniences, MFA system usually incorporates inherence factor which is biometrics, as it is something that always readily available for the user as it is part of the user or what the user does. Due to its availability at all times, it is not uncommon for biometrics to be implemented with other authentication factors. Example of this can be seen in touch dynamic with password or PIN [27], [28], passport reader with fingerprint and facial recognition, and others.

And within the biometrics, physiological biometrics are most likely to be favorable due to its convenience over that of behavioral biometrics. However, this convenience can be a double-edged sword because, as mentioned before, once physiological biometrics original information is stolen, it is difficult to reset, thus making the multi-factor authentication less secure since the false acceptance rate (FAR) of a stolen physiological biometrics can be closed to 100%.

In this dissertation, a multi-factor authentication has been proposed using rhythmic-based dynamic hand gesture named one-time rhythm. One-time rhythm, similar to OTP requires a rhythm generator to generate a rhythm for the user to shadow, an ownership factor; and an additional behavior recognition from the rhythmic gesture, an inherence factor. This means that the rhythm recognition and the behavior recognition are independent from each other in the verification process. Unlike most MFA or 2FA which uses static-static authentication or static-dynamic authentication mentioned in Chapter 2.2.2, one-time rhythm uses dynamic-dynamic authentication, i.e., both authentication methods are not always the same or changes within a short period of time. Because both

authentication methods in one-time rhythm are constantly changing, it can be challenging for impostor to steal or mimic them all.

An in-depth discussion of one-time rhythm will be discussed in Chapter 4.

2.4 What is rhythmic-based dynamic hand gesture?

Rhythmic-based dynamic hand gesture is the dynamic hand gesture – time related hand gesture, based on certain rhythm. It is considered as a behavioral biometrics. The rhythm used can be in the form of musical rhythm, pattern rhythm, and other rhythms. Rhythms can be easily memorized [29]–[31] and can be performed either consciously or unconsciously.

Rhythmic hand gesture can be obtained or captured and not limited by either the capturing the movement of the fingers, hands, and arms through different means of method. These methods used can also vary from physical movement visible to a camera device or the activity on the forearm muscles or arm muscles through electromyography technique. The actions to produce the rhythm are broad from the use of single hand to both hands, clenching the hand to moving the entire arm. As long as it can be picked up the system, it can be used rhythmic hand gesture.

In this dissertation, two different use cases of rhythmic-based dynamic hand gesture have been proposed, user-dependent rhythm and one-time rhythm. The main difference between these two use cases is that user-dependent rhythm uses rhythm created or produced by the user themselves; whereas one-time rhythm requires the user to shadow rhythm generated by a generator similar to that of one-time password (OTP).

User-dependent rhythm is similar to keystroke dynamics, speaker verification, and dynamic signature verification. It consists of a content and a style which will be considered as a whole during the verification process. On the other hand, one-time rhythm is considered as a multi-factor authentication. The rhythm used to perform is generated through a token which is considered as ownership factor, while the behavior of the gesture of performing the generated rhythm is considered as inherence factor. In one-time rhythm, the rhythm recognition and behavior recognition are independent from each other during the verification process.

2.4.1 Rhythm as static password

It may not be obvious from the sensory of human perspective but behavioral biometrics can be inconsistent from the perspective of a system. Therefore, behavioral biometrics usually contains two parts, content and style. The content can be seen from the example of characters to be typed in keystroke dynamics, text to be spoken in speaker verification, and text to be written dynamic signature. These contents can be split into two different types, fixed content and free content. Fixed content is content that has been pre-defined

and requires the user to follow the content to verify themselves. Without these pre-defined contents, the accuracy of correctly authenticating an individual using behavioral biometrics can be lower [32], [33].

As for rhythmic-based dynamic hand gesture, musical rhythm is being introduced as the content in the performance of dynamic hand gesture, which will be known as rhythmic gesture. The musical rhythm acts as a stabilizer while the performance of hand gesture that contains behavior defines the uniqueness of the user. As to why musical rhythm has been chosen is due to its simplicity in terms of performing with hand gesture. Musical rhythm can be easily familiarized and does not require special expertise to perform [34]–[36].

Although rhythm may not be viewed as static as it is time-related, the pattern of a rhythm can be viewed as static. Instead of viewing the patterns from a temporal dimension, the patterns can also be viewed from spatial dimension within a certain limited time frame. This means that as long as all the patterns from a rhythmic gesture match the intended rhythm, given that the time frame of each pattern is within a certain margin of error, then the rhythmic gesture will be deemed as the intended rhythm. Thus, the rhythms that are being used in my research can be viewed as a static password. An example of this can be seen in Figure 2.2, where the rhythm is being matched with its beat using binary signal. The orange dashed vertical line is the time frame of which each pattern should be detected. If the pattern within the given time frame matches the pattern of the intended rhythm, i.e., when there is a beat, the signal will be 1, and when there is no beat, the signal will be 0. For example, as illustrated in Figure 2.2, on an eighth beat, there is two 1's; on a sixteenth beat, there is one 1's; on a rest note, it becomes 0. When all the patterns are correctly matched, then the rhythmic gesture will be deemed as the same as the intended rhythm.

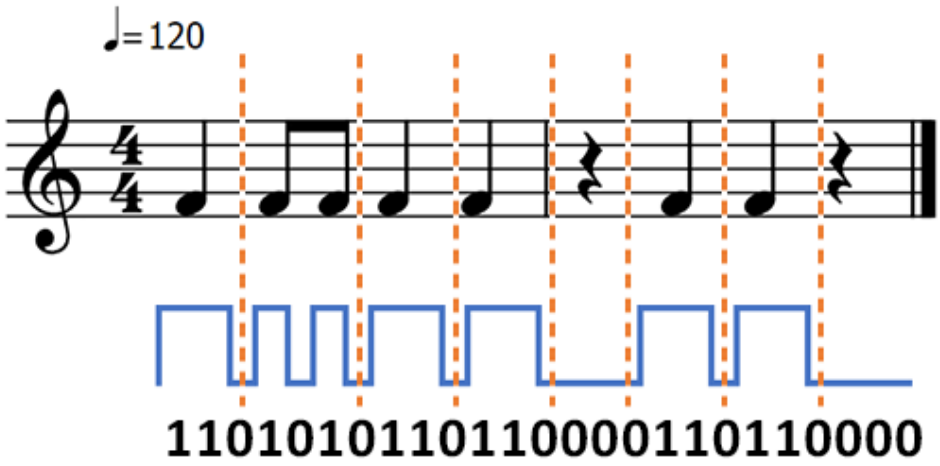


Figure 2.2: Matching pattern from rhythm to its beat in binary signal

2.4.2 Electromyography (EMG)

Electromyography (EMG) has been used to record the rhythmic hand gesture data. EMG is a technique for evaluating skeletal muscle activity, which in our case the forearm muscle as the rhythmic hand gesture is being performed. The waveform of the signal can provide information of the rhythmic gesture that has been performed by an individual. The timing and the power produced by the users while performing the rhythmic hand gesture should be able to provide behavioral information of the users.

There are two types of EMG – surface EMG (sEMG) and intramuscular EMG [37]. sEMG uses non-invasive electrodes, which can be wet or dry, that are usually placed on the surface on the skin; whereas intramuscular EMG uses invasive electrodes, usually in the form of needles or fine-wires, inserted into the muscle tissue.

The equipment used for our experiment is an sEMG wearable device known as Myo Armband [38] shown in Figure 2.2.

Figure 2.2: Myo Armband setup (top), correspondence of Myo Armband EMG sensors to their signals (bottom).

The Myo Armband consists of eight EMG sensors circulating around a flexible band. Myo Armband has the advantage over intramuscular EMG due to it being non-invasive. And since it uses dry electrodes, it has lesser chance of causing skin irritation as compared to wet electrodes [39], and requires less preparation [40]. Due to the Myo

Armband wearability, it can be easily strapped onto the forearm, and vice versa. The Myo Armband sensor is only capable of measuring muscular activity at a sampling frequency of 200 Hz, which is considered low [41], [42]. But, due to the simplicity of the gestures performed for our experiment, the sampling frequency of the Myo Armband has shown to be sufficient.

There are other techniques that can pick up rhythm from hand gesture, such as 2D camera, depth sensor, pressure sensor, and others. Table 2.1 has listed out different techniques that are capable for detecting rhythmic-based hand gesture with their availability, influencing factors, advantages, and disadvantages. Although all of the techniques mentioned have higher availability as compared to EMG currently, company such as Meta is developing EMG to interact with virtual reality [43], [44] which hints EMG to be more common in human-computer interaction (HCI). In addition, EMG has lesser external influencing factors as compared to other techniques as the sensors itself is attached directly to the body. Unlike vision-based sensor, i.e., camera, depth sensor, and radar, EMG has no line of view limitation. This can cause difficulty in distinguishing each finger when they are being stacked together or behind one another, perpendicular from the view point of the camera [45]. Although accelerometer and pressure sensor can be used similarly to EMG in terms of recognizing rhythmic-based dynamic hand gesture, they are prone to more external influencing factors such as body movement (if the user moves, the value in accelerometer changes), surrounding temperature, and gravitational artifact for accelerometer; material used, shape and size of sensors for pressure sensor. In terms of calculation, time-sequential EMG signal is two-dimensional (1D-signal and time), whereas time-sequential image (x-axis, y-axis, and time) is three-dimensional, time-sequential 3D image and accelerometer are four-dimensional (x-axis, y-axis, z-axis, and time); thus, EMG signal processing which is a 1D-signal requires less computational power as compared to that of other techniques [46].

There has been other research on hand gesture using EMG by detecting behavior from wrist movement [47] and hand shape [48]. As for rhythmic-based dynamic hand gesture, different hand gesture has been used. The hand gesture is performed by clenching the fist and relaxing the fist which is shown in Figure 2.3. The forearm muscle can be easily activated by simply clenching the fist which is shown in Figure 2.3 (b) by tightening the fingers towards the palm with force. From the rhythm point of view, this means that a beat has been successfully performed. Whereas, the forearm muscle relaxes when the fist and fingers relax which can be seen in Figure 2.3 (a). From the rhythm point of view, this means there is neither no beat or a rest note involved. This hand gesture can be seen as an advantage over the wrist movement gesture and hand shape gesture due to its little movement. This means that fist clenching gesture does not require any space to perform. Not to mention, rhythmic gesture with EMG can also be performed when the hands are busy, such as during driving. In addition to being difficult to detect by others, it can also be hidden from sight such as performing in a pocket.

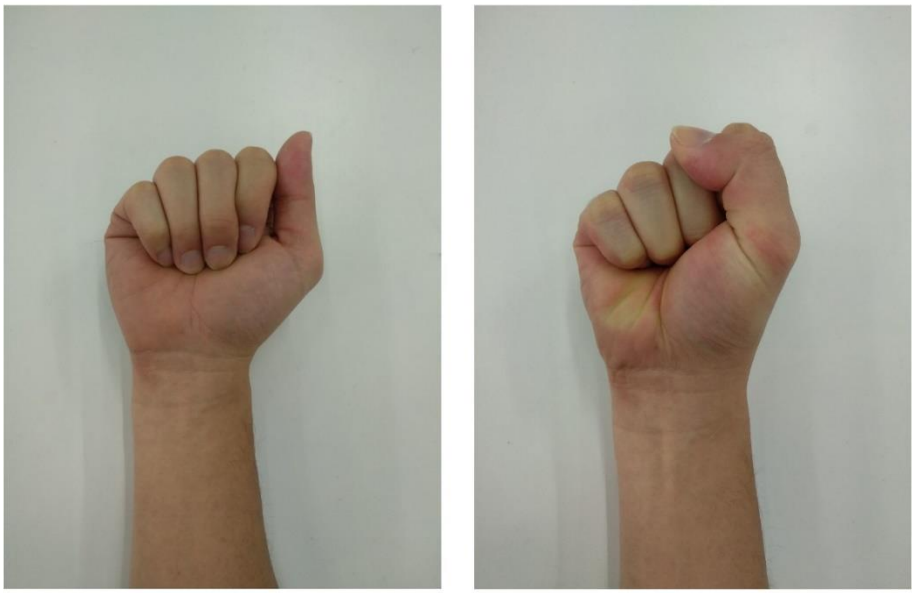
In short, the main reason EMG has been chosen to pair with rhythmic gesture are:

1. EMG has fewer external factors affecting the data

2. Signal is easy to measure
3. The combination technique can be difficult for third person to perform the measurement
4. Signal produced by the EMG exploits the inherent feature of the user

Table 2.1: Techniques available for detecting rhythmic-based dynamic hand gesture

Techniques	**Current Availability**	**Influencing Factors**	**Advantages**	**Limitations**
2D Camera	Extremely High	• Distance • Line of sight • Lighting	• Availability	• Easily obstruct • Detect other person's hand
Depth Sensor	High	• Distance • Line of sight • Lighting	• High degrees of freedom (DOF)	• Easily obstruct • Detect other person's hand
Radar [49], [50]	Medium	• Distance • Line of sight • Surrounding noise	• High DOF	• Surrounding signal noise • Detect other person's hand
Accelerometer	High	• Body movement • Temperature • Gravitational artifact [51]	• High DOF	• Movement constraint
Pressure Sensor	Medium	• Material used [52] • Shape and size of sensor [53]–[55]	• Nearly invisible	• Require hand contact • No standard sensor placement
sEMG	Low	• Anatomical factors [56] • Body movement	• Nearly invisible	• No standard electrode placement



(**a**) (**b**)

Figure 2.3: Rhythmic gesture by (**a**) relaxing the fist and (**b**) clenching the fist.

Chapter 3

Rhythmic-based dynamic hand gesture: a robust biometrics

For biometrics to be robust, it has to at least achieved most of the aforementioned desirable properties of biometrics: uniqueness, permanence, collectability, performance, and circumvention.

An experiment with various conditions has been devised to test out whether rhythmic-based dynamic hand gesture is truly robust.

3.1 Robustness of rhythmic-based dynamic hand gesture

3.1.1 Experimental setups

A total of 13 male participants, ranging between the age of 22 and 29, participated in the experiment. Participants suffering from physical disabilities, pain, and diseases causing potential neural damage, systemic illnesses, language problems, hearing or speech disorders, and mental disorders were excluded. Figure 3.1 illustrated the overall experimental conditions. The experiment was performed once every week for four weeks. In every session, the participants have been asked to perform their gesture 10 times in each scenario. In between the 10 gestures, that is, after 5 gestures, the device would be taken off and reattached onto the forearm. The refitting (taking off and then putting on again) of the device can change the position of the device on the forearm and the amount of contact of the sensors with the skin, which can affect the strength of the EMG signal. Each participant was instructed to create and perform a rhythmic hand gesture which they are more familiar with, which will be known as self-dependent rhythm. This is to ensure that the participants can easily memorize their gesture and also less variation in terms of their rhythmic tempo in the gesture [57]. The experimenter also made sure that none of the rhythm used was repeated among the participants.

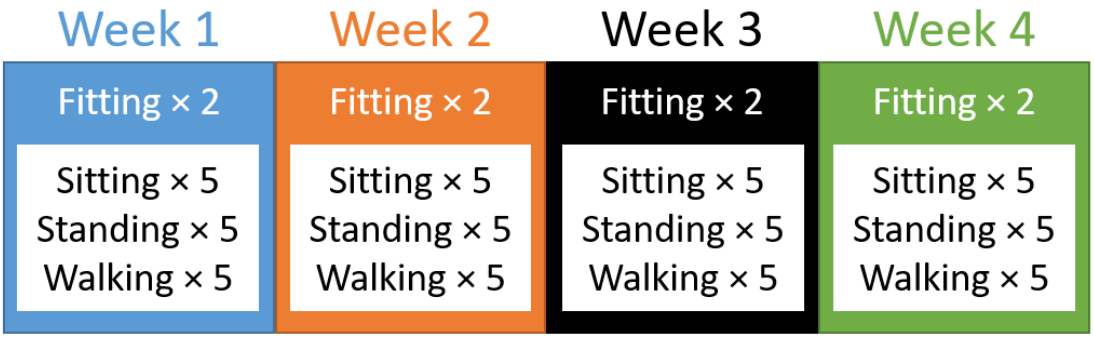


Figure 3.1: Overall experimental conditions

In this experiment, we set three different scenarios for the gesture to be performed in – sitting, standing, and walking. This is to reflect and simulate the common scenarios

of an individual interacts with their electronic device. Although sitting and standing may seem similar in terms of being stationary and not moving, as shown in Figure 3.2, people tend to bend their elbow when seated on a chair, which contracts the muscles in the forearm, as opposed to that while standing, where the arm is in a straight and relaxed posture. This may change the performance behavior of an individual. Walking, basically, will have a sudden spike in the signal due to the vibration or arm swinging while walking.

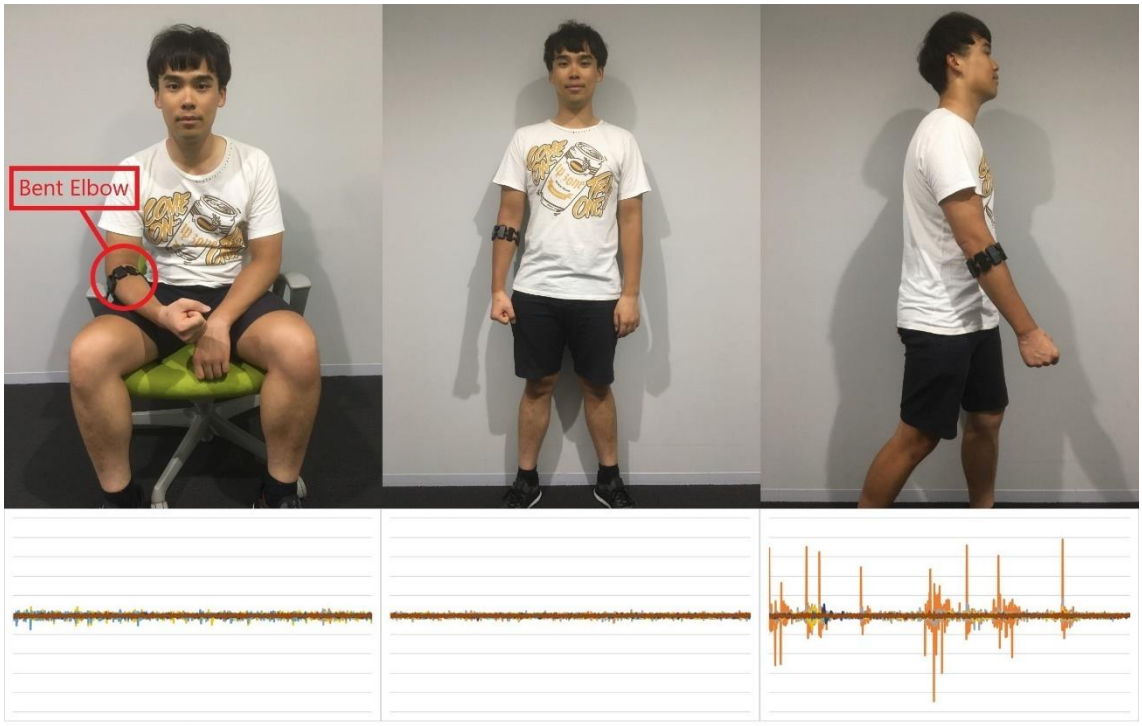


Figure 3.2: The scenarios used during the performance of the gesture:
(**a**) Sitting scenario. (**b**) Standing scenario. (**c**) Walking scenario.

Each participant was also instructed to mimic a different participant's gesture each week (excluding participant #1 who mimicked 2 different participants gestures each week due to the odd number of participants). Each participant will be mimicking 4 different participants' gestures overall, except participant #1 who mimicked 8 different participants' gestures. The genuine user visually performed their gesture in front of the impostor, as shown in Figure 3.3. This is to simulate that the impostor has clearly seen how the gesture is being performed visually.

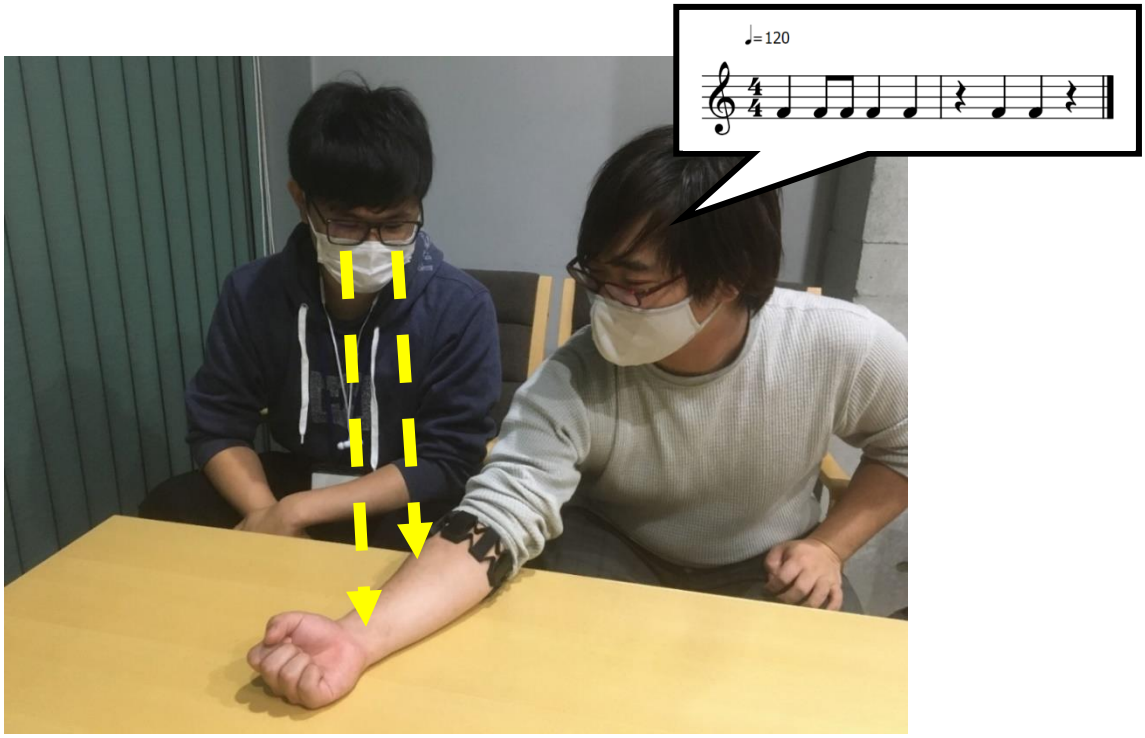


Figure 3.3: Genuine user showing their rhythmic gesture to the impostor for mimicking.

3.1.2 Data Processing

As for data processing and analysis, Python programming language, signal processing, and neural network has been used. The neural network used for our data analysis is TensorFlow Keras. Neural network has been chosen due to its ability to learn and extract features without human intervention, and later recognize new biometric samples in an unsupervised manner [58]. This means that neural network does not require predefined rules to extract specific features or representations of data [59]. In addition, neural network can be used for tasks that are non-linear [60], such as decision boundaries for non-linear data [61].

The gestures were recorded once with each fitting, which makes it about five continuous gestures in each recording. The continuous gestures have small pauses in between each other. This allows the separation of each gesture with more ease. The EMG signals were filtered using second-order Butterworth low-pass filter with 10 Hz cutoff. Although other research on EMG has been using higher sampling rate in filtering, we found out that, in our research, the applied low-pass filter has been sufficient to output the shape of the gesture, as shown in Figure 3.4. Next, resampling using linear interpolation was applied due to the imbalance in the sampling numbers in every gesture. All gestures were resampled to 600 samples. Similar to the sample size, the window size and sliding rate were set to 600 samples.

The neural network from the TensorFlow Keras Sequential model was used to predict the result. The setup of the neural network is 9 inputs, which are the 8 different channels of the EMG sensors and the original timestamp of each gesture, 10 hidden layers

with 100 nodes for each layer, and 40 epochs. The loss function used is categorical cross entropy, the optimization function used is ADAM, and the batch size used in the training phase is 1000.

The data used in training and testing have included labels indicating the participant's number. After the testing is done, the neural network will output a vector of estimation or probability consisting all the labels. This estimation ranges from 0 to 1 and sum to 1, where 0 means furthest from the label, and 1 means closest to the label. For example, training data consists of all 13 labels, i.e., all 13 participants, and when input testing data labeled 7, i.e., participants #7, the test result will output an estimation vector of the tested label to all the trained label, e.g., label 1 = 0.01, label 2 = 0.0, …, label 7 = 0.91, …, label 13 = 0.02. From the example, it can be said that the tested data that has been labeled 7 has the highest estimation of 0.91 similarity to labeled 7 of the trained data, thus the test data belongs to participant #7 which is the correct match.

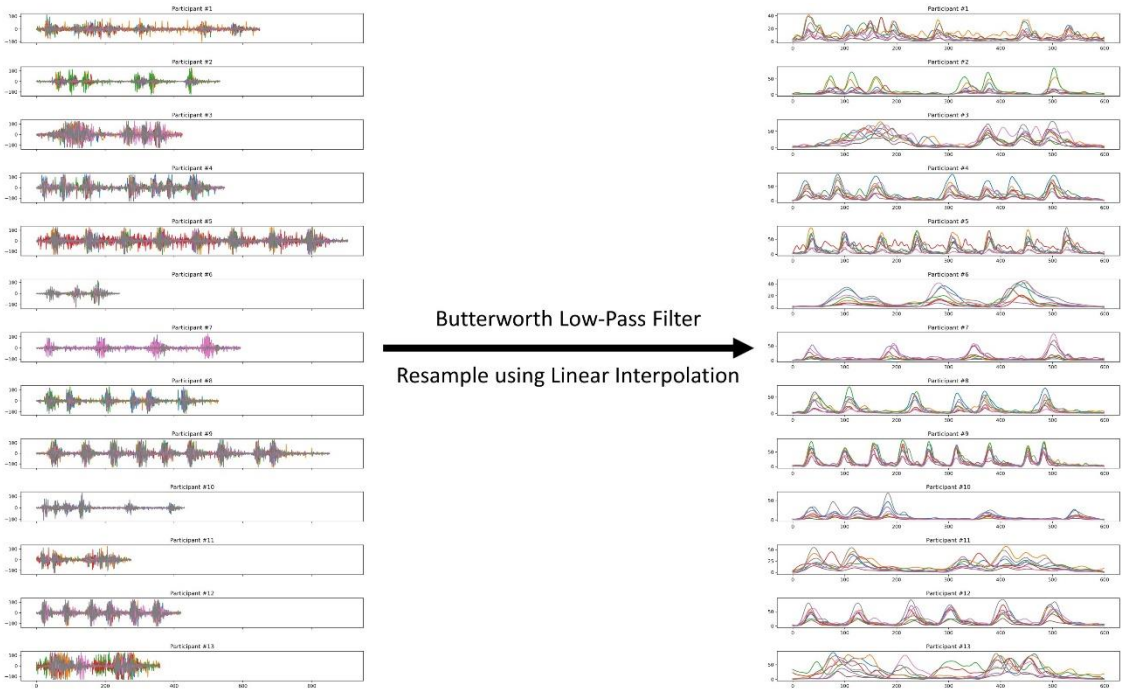


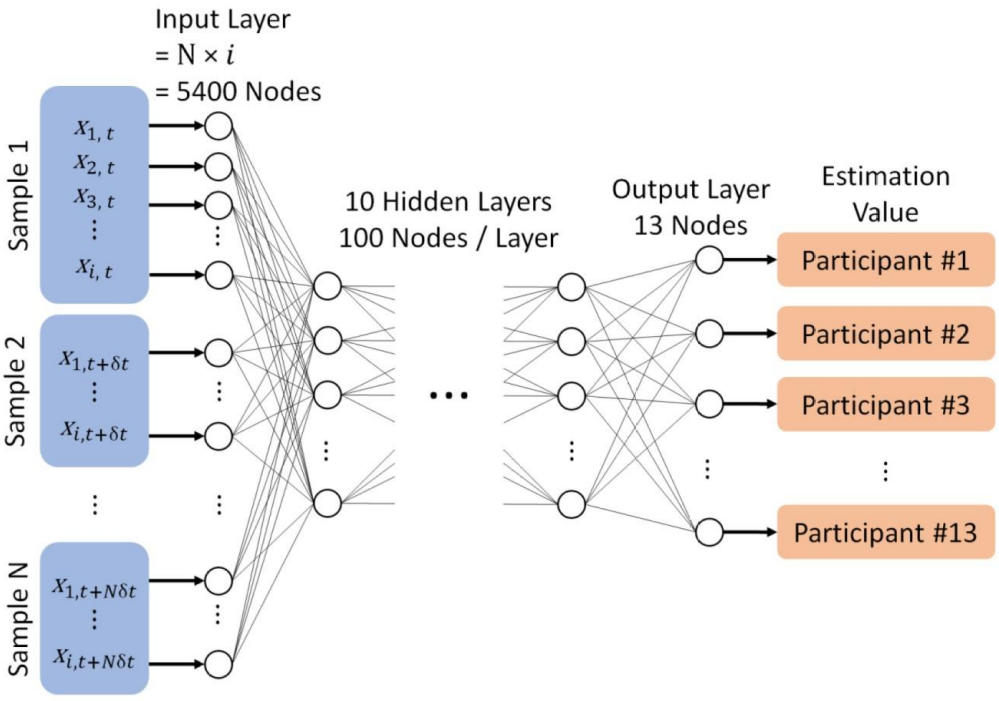
Figure 3.4: Samples of the raw signal (**left**) and processed signal (**right**) after being filtered using Butterworth low-pass filter and then resampled using linear interpolation.

We consider a set of weights as coefficient values of a digital filter. It means that a well-learned neural network using our data set is expected to have coefficient values of finite impulse response (FIR) filter in order to judge "who is who". The benefit of the time-order array design is to obtain time-space features in the training phase, as well as

easy to achieve real-time processing. Basic design of data structure, given the obtained EMG signal, is arranged in time-order, as shown in the blue box (left side) of Figure 3.4.

Because the experiment was done in multiple sessions, the fitting of the EMG device, which affects the EMG sensors' location, were always different. To overcome the inconsistent physical location of the EMG sensors that occurs after the refitting the device, I proposed a virtual rotation in algorithm. This is done by rotating the EMG input of the neural network, as shown in Figure 3.5. The virtual rotation also serves as extra training data. This rotation can be done due to position of the electrodes on the Myo Armband are arrange in a circular pattern and are symmetry to each other. Because of this circular pattern and symmetricity, the signals can be positioned different without changing the overall shape of the signal. However, due to how the neural network learn the data which the arrangement of the electrodes is taken into account, this symmetry rotation can help prevent overfitting when there is only a single position being used in the training data.

Finally, the results will be evaluated with false acceptance rate (FAR), false rejection rate (FRR), area under the curve (AUC) of receiver operating characteristics (ROC) curve, and equal error rate (EER).



i = {EMG 1, EMG 2, EMG 3, EMG 4, EMG 5, EMG 6, EMG 7, EMG 8, Original Time Stamp}
N = 600

Figure 3.4: The neural network layers and computational data.

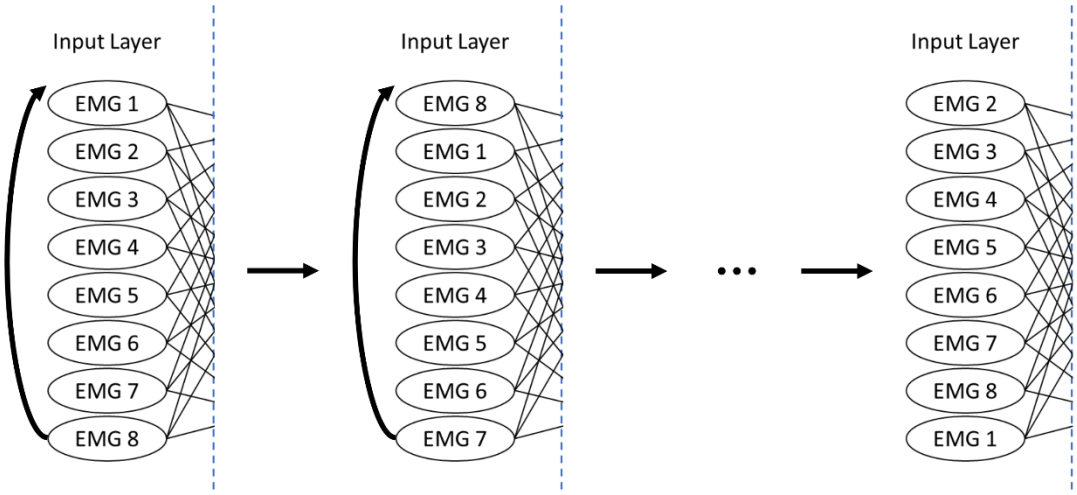


Figure 3.5: Virtual rotation of the 8 channels of electromyography (EMG) sensors in the neural network.

3.1.3 Experimental conditions

To test out the robustness of rhythmic-based dynamic hand gesture with EMG, different conditions have been analyzed. The different conditions are:

1. Different scenarios in training and testing,
2. Single fitting multiple trials versus multiple fitting with one trial each,
3. Comparisons between no rotation and virtual rotation over different period of time,
4. Mimicking other participants gestures.

The speed of validation on each gesture data is up to 2 ms. Though it is slower than fingerprint validation, which can be validated in less than 100 μs [35,36], it can be very similar to face recognition, with recognition speed of 2.4 ms [37]. In normal use, it is fast enough to be indiscernible by users.

Condition 1:

The purpose of this condition is to evaluate whether different scenarios, i.e., sitting, standing, and walking can have significant effect on the performance of rhythmic-based dynamic hand gesture using EMG. The methodology of evaluating this is by splitting the data into different scenarios in training and testing, where the training and testing should not have the same scenarios. Every training and testing data set for each participant consists of 1 scenario × 8 fittings × 5 trials:

1-1. Training = Sitting scenario; Testing = Standing scenario.

1-2. Training = Sitting scenario; Testing = Walking scenario.

1-3. Training = Standing scenario; Testing = Sitting scenario.

1-4. Training = Standing scenario; Testing = Walking scenario.

1-5. Training = Walking scenario; Testing = Sitting scenario.

1-6. Training = Walking scenario; Testing = Standing scenario.

Figure 3.6 illustrates the summarized data set of Condition 1, where the three scenarios have been separated into three different enrollment set. Each enrollment set is then trained in different neural network models which are independent from each other. As for the probe set, scenarios different from the enrollment set have been also been segregated into separated probe sets within the remaining scenarios. For example, when sitting scenario is used for enrollment, standing and walking are both split into two different set of probe sets. Each of the probe set is then tested and evaluated with the neural network model separately. The evaluation of FAR, FRR, AUC, and EER are summarized in Table 3.1, Condition 1-1 to 1-6, of Chapter 3.1.3.

Condition 2:

This experiment is to test whether refitting the sensors affect the results in authenticating the user. The single fitting with multiple trials training data set consists of 3 scenarios × 1 fitting × 5 trials, i.e., participant fits the Myo Armband once and performs the gesture 5 times without changing or repositioning the sensors; whereas the multiple fittings with one trial each training data set consists of 3 scenarios × 5 fittings × 1 trial, i.e., participant puts on the Myo Armband and performs the gesture once and then refits or repositions the Myo Armband by taking it off and putting it back on; this repetition is done 5 times:

2-1. Training = 3 scenarios × 1 fitting × 5 trials.

2-2. Training = 3 scenarios × 5 fittings × 1 trial.

Both conditions have the same amount of training data. Both testing data consist of the same 3 scenarios × 2 fittings × 5 trials, i.e., 2 fittings from week 4. Figure 3.7 illustrates the summarized data set of Condition 2, where one-fitting and multiple-fitting are split into two different enrollment set. Each enrollment set is then trained in different neural network models which are independently from each other. As for the probe set, two different fittings that are not in the enrollment sets have been used. The probe set for testing one-time fitting model and multiple fitting model is the same for fair comparison. The probe set is been tested and evaluated independently among the one-fitting model and multiple-fitting model. The evaluation of FAR, FRR, AUC, and EER are summarized in Table 3.1, Condition 2-1 to 2-2, of Chapter 3.1.3.

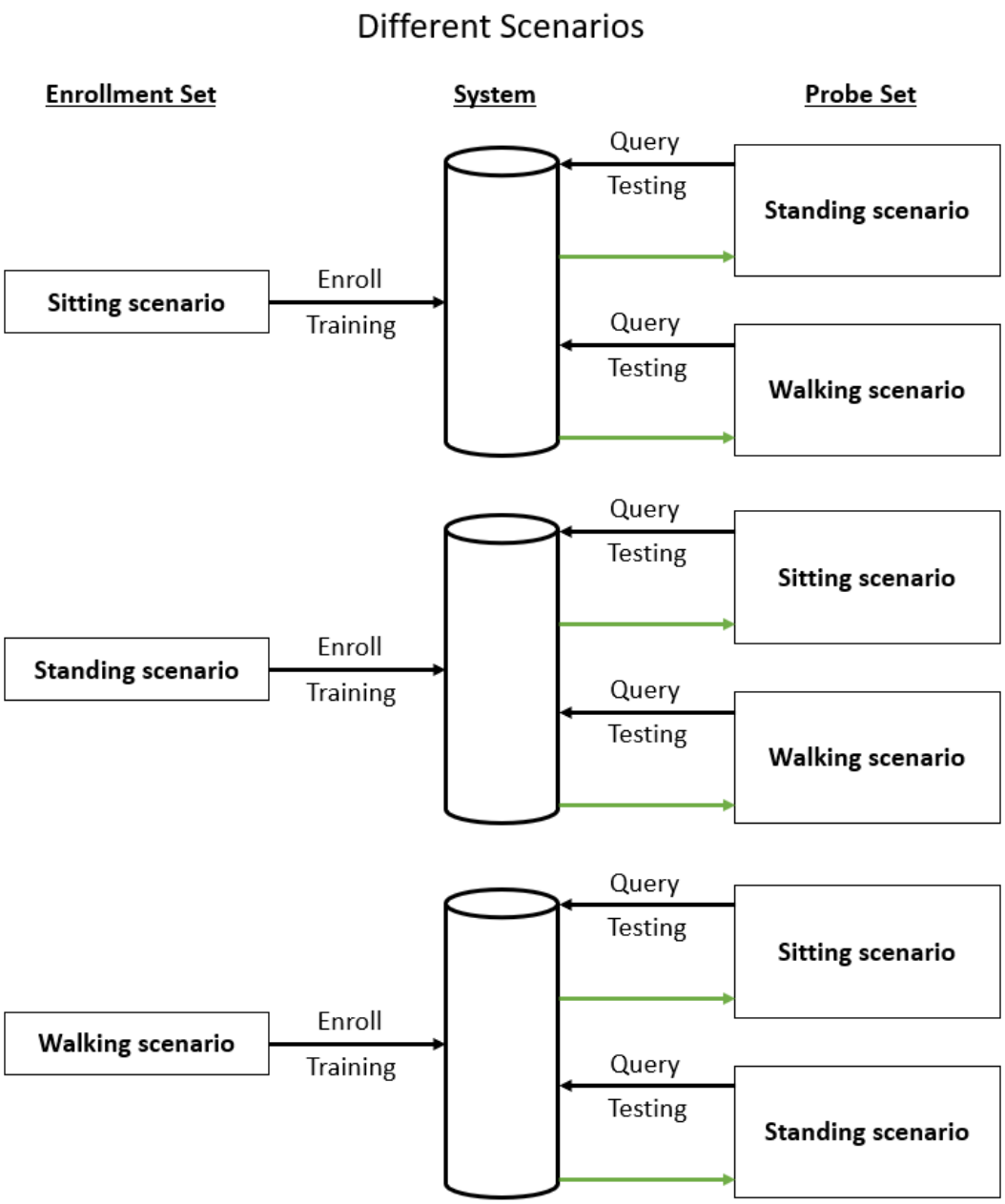


Figure 3.6: Different scenarios condition

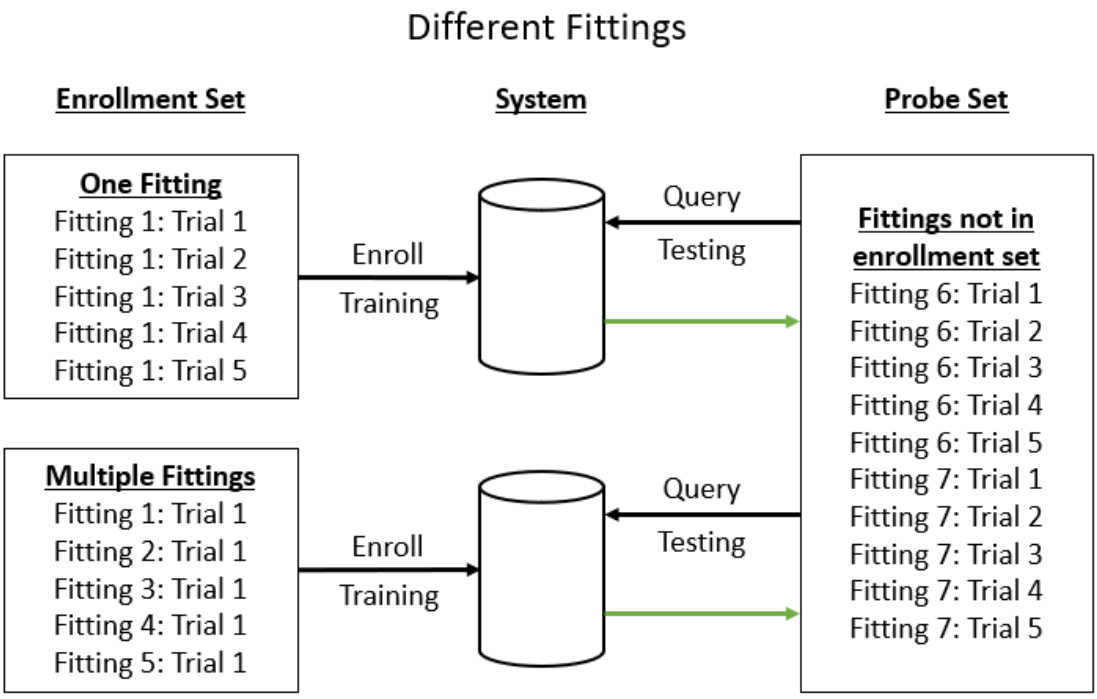


Figure 3.7: Different fittings condition

Condition 3:

This condition is to evaluate how time period affects rhythmic-based dynamic hand gesture, in addition to the effect of virtually rotating EMG data. The time period is referred to the amount of time period that has been enrolled into training the neural network model, i.e., one-week data or multiple-week data. The idea of virtually rotating EMG data is to produce more variation of data as if the EMG device has been refitted. The data for this condition is as follows:

3-1. Training = 3 scenarios × 2 fittings (1 week) × 5 trials,

Testing = 3 scenarios × 2 fittings (1 week) × 5 trials.

3-2. Same as 3-1 but with virtual rotation

3-3. Training = 3 scenarios × 6 fittings (3 week) × 5 trials,

Testing = 3 scenarios × 2 fittings (1 week) × 5 trials.

3-4. Same as 3-3 but with virtual rotations

Every week, the rhythmic gesture was performed with 2 different fittings. For condition 3-1 and 3-2, week 1 data has been used to compare with week 2, week 3, and week 4; e.g., week 1 – week 2, week 1 – week 3, week 1 – week 4, with 2-fold cross-validation each. For condition 3-3 and 3-4, 3 weeks data has

been used as training and 1 week data has been used as testing with 4-fold cross-validation. Figure 3.8 (a) illustrate the summarized data set of Condition 3-1 and Condition 3-3, where one-week data and multiple-week data are split into two different enrollment set. Each enrollment set are then trained in different neural network models which are independent from each other. The probe set consists of one-week data that is not in the enrollment set, i.e., enrollment set consists of week 1, 2, and 3, probe set consists of week 4. The probe set is then tested and evaluated independently among the one-week model and multiple-week model.

Figure 3.8 (b) illustrate the summarized data set of Condition 3-2 and Condition 3-4. The difference is that the enrollment set have been implemented with virtual rotation, i.e., each data has been enrolled with 8 different rotation patterns. The evaluation FAR, FRR, AUC, and EER are summarized in Table 3.1, Condition 3-1 to 3-4, of Chapter 3.1.3.

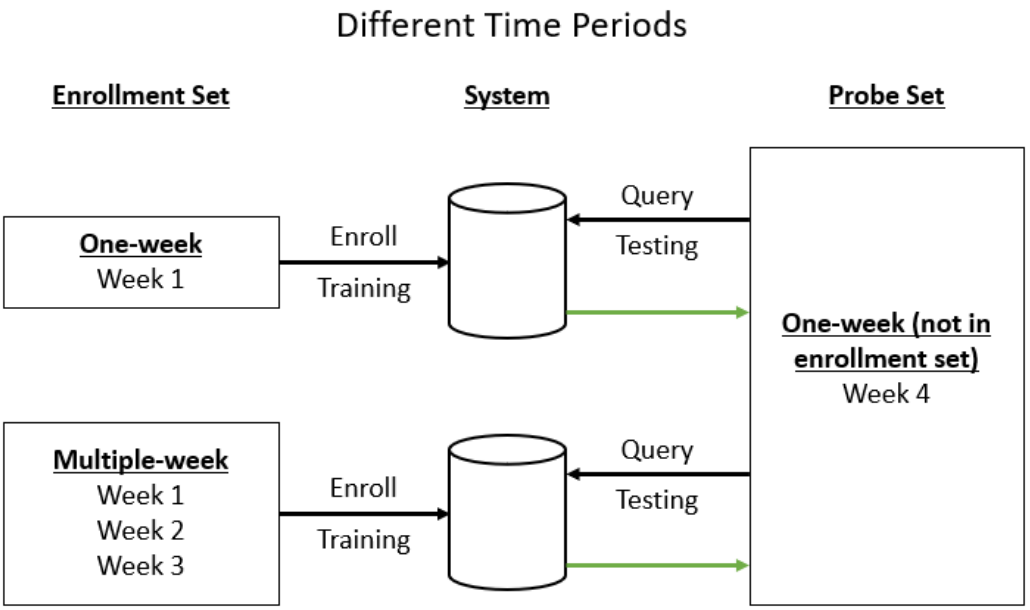


Figure 3.8 (a): Different time periods condition

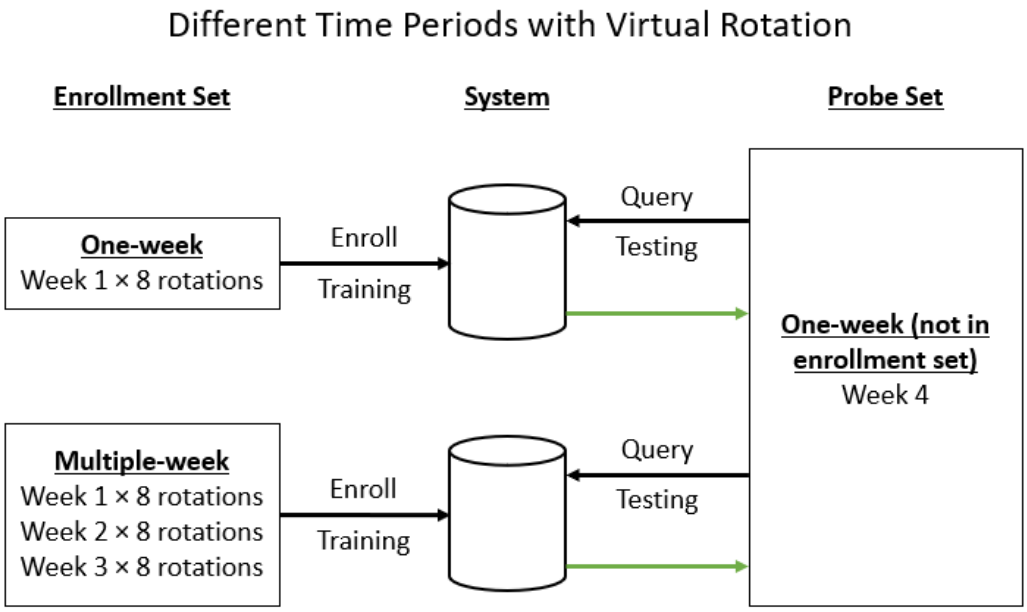


Figure 3.8 (b): Different time periods condition with virtual rotation

Condition 4:

The analysis on this condition is different from all the aforementioned conditions. Instead of authenticating genuine users, this experiment is to identify how many impostors were able to successfully authenticate as genuine users using their mimicking data. We will be using FAR – impostors being falsely authenticated as genuine users to analyze the results. The training data set consists of genuine user's 3 scenarios × 8 fittings × 5 trials and each impostor's 4 participants' gesture × 3 scenarios × 1 fitting × 5 trials (This excludes participant #1 where there are 8 other participants' gestures instead of 4.). Each genuine user's data was mimicked by the impostor for 4 times. The testing data set consists of impostor's 4 different participants' gestures × 1 fitting × 5 gestures (This excludes participant #1 where there are 8 other participants' gestures instead of 4). Both datasets were applied with 4-fold cross validation:

4-1. Training = All genuine user data + 3 of impostor mimic data

Testing = 1 impostor mimic data different from training data

4-2. Same as 4-1 but with virtual rotation

The reason impostors' data were added into the training data is due to the arrangement of the neural network used which is an identification-oriented neural network. Without impostor's data in the training data, the result will end up with only pointing to the genuine user. Figure 3.9 (a) illustrates the summarized data set of Condition 4-1, while Figure 3.9 (b) illustrates the summarized data set of

Condition 4-2, where the enrollment set's data has been implemented with virtual rotation. The enrollment data consists of all the genuine user data and three of the impostor's mimicking data with five trials each. The enrollment set is then trained in the neural network model. The probe set consist of the impostor's mimicking data that is not in the enrollment set. The probe set is then tested and evaluated on the neural network model. The FAR is summarized in Table 3.1, Condition 4-1 to 4-2, of Chapter 3.1.3.

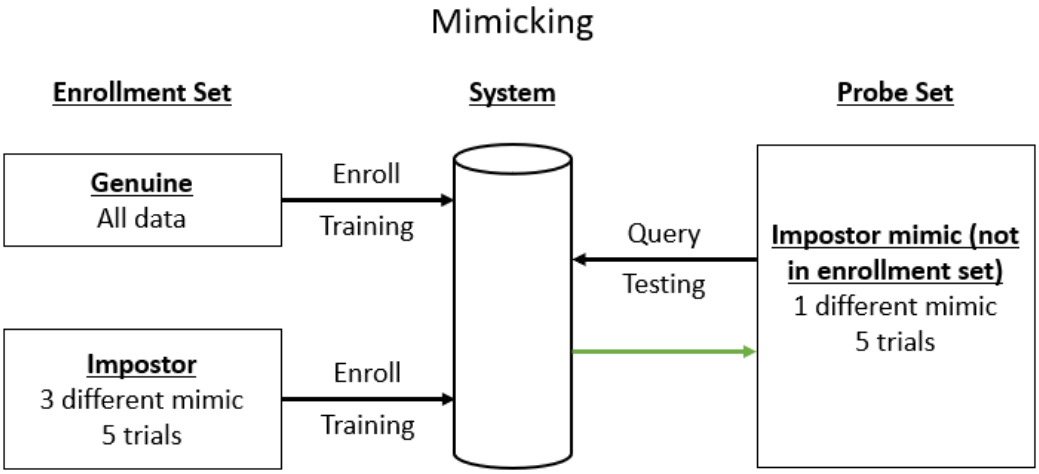


Figure 3.9 (a): Mimicking condition

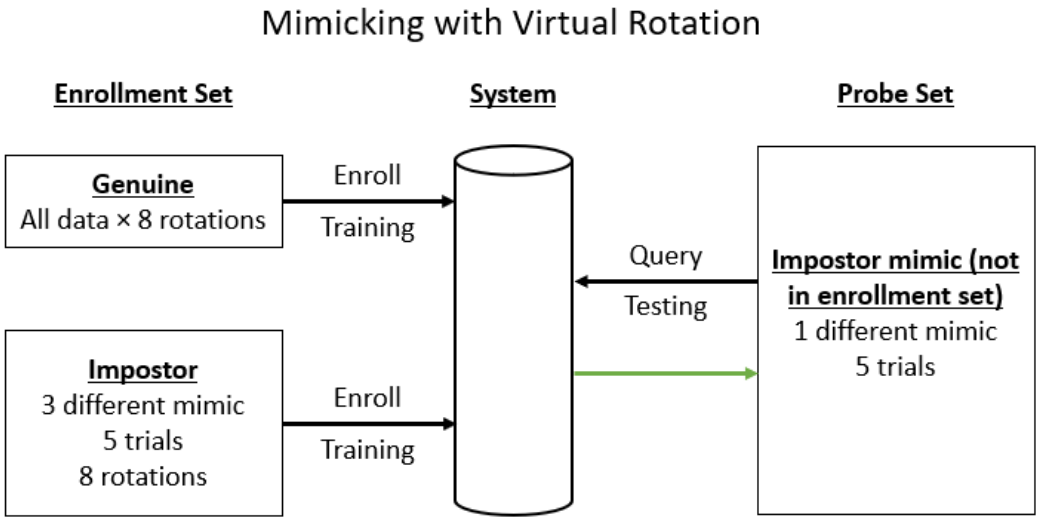


Figure 3.9 (b): Mimicking condition with virtual rotation

3.1.4 Biometric characteristics of rhythmic-based hand gesture

The first desire property of biometrics, universality can be easily confirmed as almost everyone with a hand will be able to produce rhythmic-based dynamic hand gesture. Of

course, with the exception with certain illnesses and disorders that restrict or prohibit a person from performing the hand gesture.

To collect the rhythmic gesture, there are many methodologies that can be used to acquire, such as but not limited to EMG, camera, radar, and pressure sensor. In this dissertation, the focus is on wearable EMG as it easier to setup, and even minute movement from the muscles can be easily detected. In addition to being hands-free, non-invasion, and wireless. Thus, collectability has been achieved.

Permanence has been realized through the experiment with the confirmation on performance as well. The conditions from the experiment that confirmed permanence and performance are the different scenarios – sit, stand, and walk; as well as the different session that has been split into 4 different week times. Since permanence is to test how resistance biometrics is to changes that can be either from surrounding influences or time influences. Due to how accurate can the system correctly recognized who performed the rhythmic gesture in the different scenarios and time period, this also testify how well the biometric system has performed.

In Table 3.1, it can be seen that different scenarios can have slight influence in performance, especially while in motion (walking), due to the difference in behavior and noise produced. But, even with the differences, EER is still as low as 4.62%. This proves that intrapersonal difference is smaller than that of interpersonal difference in terms of different scenarios. With variation of scenarios in the training data, the difference can be less significant. This can be seen in Condition 3-3 of Table 3.1, where all scenarios are trained together.

In real-life situation, the EMG sensor is expected to be taken off and worn back on from time to time. This has been tested and evaluated as can be seen in Condition 2-1 and 2-2 of Table 3.1 which includes single fitting with multiple trials (worn once and perform multiple trials) and multiple fittings with one trial each (taken off and worn back on multiple times and perform one trial each time being worn). The EER of multiple fitting is 4.85% which has higher performance than that of single fitting which has an EER of 10.75%. Even though both training data have the same amount of data, the multiple fitting with one trial each yields higher results than single fitting with multiple trials in one fitting. This shows that variations in training data is key to performance. It is also worth mentioning that the multiple fittings also consist of 3 different weeks data in the training data, as each week only consist of 2 fittings.

In Condition 3-1 and 3-3, training data with only one period of time and training with multiple periods of time have been evaluated. In Condition 3-1, the training data only uses one week of data which is the first week's data. This resulted in high FRR of (52.07±10.93) %. In Condition 3-3, the training data has a combination of 3 different weeks' data which resulted in a lower FRR of (7.76±3.06) %. This indicates that rhythmic-based dynamic hand gesture requires frequent update by adding new training data. Based on Condition 3-1 and 3-2, and Condition 2-2, it can be concluded that rhythmic-based dynamic hand gesture has permanence in different time period as long as the training data is updated or added with new training data from different period of time.

In Condition 3-2 and 3-4, virtual rotation has been introduced. This can be especially useful when there is a limitation in collecting or updating training data. As can be seen in Condition 3-2, when virtual rotation has been introduced, the FRR drops from (52.07±10.93) % in Condition 3-1 to (38.47±2.42) % in Condition 3-2. But as training data is being updated as illustrated in Condition 3-3 and Condition 3-4, the virtual rotation does not provide any advantages.

In biometrics, how a system can easily tricked by fraudulent techniques is known as circumvention. It is also one of the most important properties in biometric authentication as biometric authentication is implemented to prevent other individuals from accessing the system in the first place. To test out the circumvention of rhythmic-based dynamic hand gesture, every participant was requested to mimic another participants' rhythmic gesture shown in Figure 3.2 to find out whether the system is able to detect the impostors trying to mimic genuine user's gesture. Due to the unchangeable nature of static biometrics, once it is stolen or forged, it cannot be reset, which will lead to a high FAR [1], [15], [16], and higher FAR indicates ease of circumvention which is not desirable. But, in dynamic biometrics, even when an impostor tries to mimic another individual's gesture, the system is able to differentiate whether the gesture belongs to the genuine user or the impostor due to the difference in behavior. This has been drawn out in the result shown in Condition 4-1 of Table 3.1. Rhythmic-based dynamic hand gesture has an FAR of as low as (10.38±1.79) % when no virtual rotation is implemented in the training data. It outperforms that of stolen or forged static biometrics, where forged fingerprint has a FAR of more than 67% [1]. In addition, the user has the choice to change their behavior or gesture even after their data has been successfully compromised, unlike static biometrics, which are unchangeable.

Condition 4-2 is the implementation of virtual rotation in conjunction with Condition 4-1. This is to test out whether virtual rotation affects the FAR because in theory, virtual rotation lowers the threshold of acceptance due to the increase of variations in the training data. From Condition 4-1 and 4-2, it can be seen that the FAR increases from (10.38±1.79) % to (15.45±3.65) %. With the results from Condition 3-2, 3-4, and 4-2, it is advice to only use virtual rotation when there are limited training data and should be removed after sufficient training data have been acquired.

With all these results, it can also be concluded that rhythmic-based dynamic hand gesture is also unique. The system is able to recognize genuine user's rhythmic gesture at EER of as low as 1.32% and also a low FAR when other individuals try to mimic the genuine user's rhythmic gesture.

3.2 Comparison with other biometrics

First, the discussion on the influencing factors that can affect the performance of the behavioral biometric modalities. Different modalities have different influencing factors and can be difficult to evaluate which influencing factor is more severe than the others. However, most biometrics modalities mentioned have external influencing factors, such

as ambient noise, tool used, terrain, temperature, and others. As for the proposed rhythmic hand gesture using EMG, the influencing factors are mostly from the anatomical factors, which is relating to bodily structure. This means that the proposed biometric modality can be used at anywhere and anytime without much of a concern of external or environmental changes. And although motion artifact can affect EMG which has also been mentioned in Chapter 3.1.4, the result shows that the performance remains high.

It can be difficult to determine which biometric modality is the best but instead depending on use case. Each biometric modality mentioned has their own advantages such as low cost, high degree of freedom (DOF) which increases features, conventional input method, and others. Although the proposed biometric modality is currently not that common in the market, it can be advantageous when it comes to extra security as it is nearly invisible during performance, in addition to the ability of hidden from sight, such as performing in pocket.

As for limitations, the proposed biometric modality has no standard in the electrode placement. It means that every time when the electrode is placed differently, it can affect the performance. This can be seen in Condition 2-1 of Chapter 3.1.3 and Chapter 3.1.4. When there is only one fitting (one position of electrodes) being used in training data, the performance of the proposed biometric modality drops significantly when the position of electrodes changes. Therefore, it is recommended to include different fittings (different position of electrodes) have to be included in the training data to increase the performance which can be seen in Condition 2-2 of Chapter 3.1.3 and Chapter 3.1.4.

Table 3.2 draws out the comparisons between different behavioral biometrics, that includes speaker verification, signature verification, gait, keystroke dynamics, keystroke sound, electroencephalogram (EEG) biometrics, acceleration-based hand gesture, and the proposed rhythmic hand gesture. All the behavioral biometrics are selected due to the similarity in terms of using behavior as authentication. EEG biometrics which detect the electrical activities of the brain can be seen similar to that of the proposed biometric modality. EEG biometrics usually requires the user to perform a pre-defined task and record the EEG signal while the task is being performed [62], similar to that of EMG signal being recorded while gesture is being performed. Both EEG and EMG also uses signal processing to extract subject-dependent features that will be used for biometric authentication. The performance metric shows in Table 3.2 is the EER of the biometric modality. Although this can be difficult to justify which biometric modality has the best performance due to the difference in data, the proposed biometric modality is relatively high as compared to other biometric modalities which has EER of as low as 1.3%.

Convenience discusses how easy can the biometrics modalities can be used. High convenience indicates that the biometrics modalities are easy to use and highly available, such as speaker verification, signature verification, and keystroke dynamics. The proposed biometric modality is medium in convenience as it requires the user to wear the device and perform the gesture for a few seconds. Also, the current availability is not as common as microphone or keyboard. But it is more convenient than that of gait and EEG biometrics, which require more preparations, such as space to perform, and electrodes setup respectively.

Reliability and security can be defined by the FRR and FAR respectively from the EER as the EER is the predetermined threshold for the biometric system's FAR and FRR. But as can be seen from Table 3.1, the FAR and FRR can range from 0.1% to 20% depending on the database and methodology used, which can be difficult to compare the reliability and security quantitatively. Therefore, besides using FAR and FRR, influencing factors, advantages, and limitations have also been used to evaluate the reliability and security of the biometric modalities in Table 3.1. Speaker verification and keystroke sound can be reliable in terms of picking up the speaker's voice or typing sound but at the same time it can also be easily interfered by the surrounding noise, making their reliability medium. Whereas signature verification depends on the tools to write and to be written on, and can vary significantly within a short period of time [63], which can great affect the FRR, making its reliability low. Gait can be easily affected by environment factors such as terrain, weather, and others; and also, intra-personal factors such as fatigue, making its reliability low. Similarly, EEG biometrics can be affected explicitly by environmental factors that causes body or eye movement which affects the EEG signal; and also, intra-personal factors such as mental state, making its reliability low. For keystroke dynamics, it has high reliability due to the direct contact with the input device and feedback produced during typing. As for the proposed biometric modality, it has a high reliability similar to that of keystroke dynamics due to the direct contact of the sensor to the skin making it less likely to be interfered or obstructed, and also the feedback produced when performing the gesture making sure that the gesture has been performed correctly.

As for security, speaker verification, keystroke sound, and signature verification have low security due to the original biometrics can be easily recorded or copied, and replayed. Although gait, keystroke dynamics, acceleration-based hand gesture can be physically captured or recorded, they require intensive practice for the behavior to be as similar as the genuine user, making them more secure than that of speaker verification and keystroke sound. As for EEG biometrics and proposed rhythmic hand gesture using EMG, they have high security due to them being not visible physically or can be hidden from sight during performance, which reduces the chance of impostors stealing the biometrics. Even though only the physical or sensor level security have been used in the comparison, it can also be worth mentioning that from the system level, cancellable biometrics can be implemented to any of the biometrics modalities in the comparison to improve the security.

3.3 Summary

The intention of this chapter is to examine and evaluate whether rhythmic-based dynamic hand gesture using EMG is a robust biometric modality. The rhythmic-based dynamic hand gesture that has been used to test the robustness is user-dependent rhythm, where the rhythmic gesture depends on the rhythm pre-defined by the user. To be robust, the biometric modality has to be reliable and secure, in addition convenient. Reliability depends on how well the performance and the permanence of the biometric modality is.

This can be measured with the EER of the biometric modality which can be as low as 1.32%. As for the security, this can be measure as to how easy impostors can be accepted by the system as a genuine user from FAR. In the mimicking scenario, the FAR is as low as 10.38% which is lower than that of FAR of physiological biometrics. Moreover, the rhythmic gesture paired with EMG has an advantage of being nearly invisible, can be hidden from sight, and due to the direct contact of the EMG sensor to the skin, it can be difficult to be stolen physically. Finally, when compared with other behavioral biometrics, the EER seems to be within the range of most behavioral biometrics. Also, rhythmic gesture with EMG has fewer external factors as most factors occurred within the human due to the direct contact of EMG to the body.

Table 3.1: Summary results of the different scenarios, different fittings, virtual rotation, and mimicking.

Condition	Training (Scenarios × Fittings × Trials)	Testing	Rotation	Cross Validation	FAR (%)	FRR (%)	AUC (%)	EER (%)
1-1	Sit × 8 × 5	Stand × 8 × 5	No	N/A	0.51	6.22	99.84	**1.32**
1-2	Sit × 8 × 5	Walk × 8 × 5	No	N/A	0.96	11.54	99.26	4.04
1-3	Stand × 8 × 5	Sit × 8 × 5	No	N/A	0.79	9.42	99.56	3.35
1-4	Stand × 8 × 5	Walk × 8 × 5	No	N/A	1.01	12.12	98.97	4.23
1-5	Walk × 8 × 5	Sit × 8 × 5	No	N/A	1.07	12.88	99.24	4.62
1-6	Walk × 8 × 5	Stand × 8 × 5	No	N/A	0.87	10.44	99.24	3.67
2-1	3 × 1 × 5	3 × 2 × 5	No	N/A	1.93	23.08	96.09	10.75
2-2	3 × 5 × 1	3 × 2 × 5	No	N/A	1.05	12.56	99.07	**4.85**
3-1	3 × 2 × 5	3 × 2 × 5	No	3 × 2	4.34 ± 0.91	52.07 ± 10.93	96.52 ± 0.97	9.54 ± 1.47
3-2	3 × 2 × 5	3 × 2 × 5	Yes	3 × 2	3.21 ± 0.20	38.47 ± 2.42	97.60 ± 0.91	7.13 ± 1.57
3-3	3 × 6 × 5	3 × 2 × 5	No	4	0.65 ± 0.26	7.76 ± 3.06	99.52 ± 0.27	3.20 ± 0.95
3-4	3 × 6 × 5	3 × 2 × 5	Yes	4	0.66 ± 0.31	7.86 ± 3.70	99.13 ± 0.86	3.22 ± 1.64
	Genuine + Impostor	**Mimic**						
4-1	3 × 8 × 5 + 3 × 1 × 5	1 × 1 × 5	No	4	**10.38 ± 1.79**	N/A	N/A	N/A
4-2	3 × 8 × 5 + 3 × 1 × 5	1 × 1 × 5	Yes	4	15.45 ± 3.65	N/A	N/A	N/A

Table 3.2: Comparison of behavioral biometrics with proposed rhythmic gesture

Behavioral Biometrics	Influencing Factors	Advantages	Limitations	Performance (%)	Convenience	Reliability	Security
Speaker verification [64]–[67]	• Ambient noise • Language and accent	• Different language and accent can be difficult to mimic	• Cannot be used in noisy place • Can be language dependent	• EER: 0.86 ~ 20	• High	• Medium	• Low (can be recorded)
Signature verification [68]–[75]	• Writing tool • Medium written on	• Low cost	• Can vary a lot depending on tools	• EER: 0.1 ~ 20	• High	• Low	• Low (can be copied)
Gait [76]–[79]	• Terrain • Clothing (camera-based)	• High degree of freedom	• Require large area	• EER: 2.2 ~ 20	• Low	• Low	• Medium (able to capture but require training)
Keystroke dynamics [80]–[83]	• Typing tools • Key layout	• Can be used with convention keyboard or touch screen	• Layout of keys or text to be typed can affect performance	• EER: 0.25 ~ 17	• High	• High	• Medium (able to capture but require training)
Keystroke sound [84], [85]	• Typing tools • Key layout • Ambient noise	• Uses conventional keyboard	• Dependent on the keyboard • Changes in text can affect performance	• EER: 4 ~ 20	• Medium	• Medium	• Low (can be recorded)

EEG biometrics	• Mental state • Eye movement	• Unforgeability	• Low spatial resolution (i.e., requires many electrodes)	• EER: 0.39 ~ 35	• Low	• Medium	• High (cannot be seen)
Acceleration-based hand gesture [6], [86]–[89]	• Motion artifact • Temperature • Gravitational artifact [51]	• High degree of freedom	• Movement constraint	• EER: 2 ~ 6	• Medium	• Medium	• Medium (able to capture but require training)
Proposed rhythmic hand gesture using EMG	• Anatomical factors [56] • Motion artifact	• Can be hidden from sight	• No standard electrode placement	• EER: 1.3 ~ 10	• Medium	• High	• High (can be hidden)

Chapter 4

One-time rhythm

4.1 What is one-time rhythm?

One-time rhythm is a technique similar to OTP where different rhythms will be generated when authentication is required. The user will be asked to perform the generated rhythm and the system will compare whether the rhythmic gesture matches the generated rhythm. But one distinct different from OTP is that, one-time rhythm also implements behavior recognition into the rhythmic gesture. This means that even if the rhythmic gesture matches the generated rhythm, if the behavior does not match to that of the intended user, the authentication system will deem it as an impostor.

Since OTP is also known as dynamic password [13], [14], meaning it changes every time when a user request for a password. Since this technique is also brought towards one-time rhythm, the later part of one-time rhythm is behavioral recognition, this makes one-time rhythm a dynamic × dynamic authentication. This means that one-time rhythm in theory should be more difficult to break into by impostors.

Figure 4.1 illustrated the flowchart of both OTP and one-time rhythm. Both processes are similar except one-time rhythm includes an extra step which is behavior recognition (check behavior). Although behavior recognition is an extra step in the process, it does not require any extra input or action from the user as the behavior is always incorporated with the rhythmic gesture.

An experiment has been designed to test out how well one-time rhythm works in terms of rhythm recognition and behavior recognition. To do so, first rhythm has to be generated. Instead of auto-generated rhythms, 8 different rhythms have been devised to test out how different beats, tempo, and length of rhythms affect rhythm recognition. These rhythms can be of either visual cues or auditory cues.

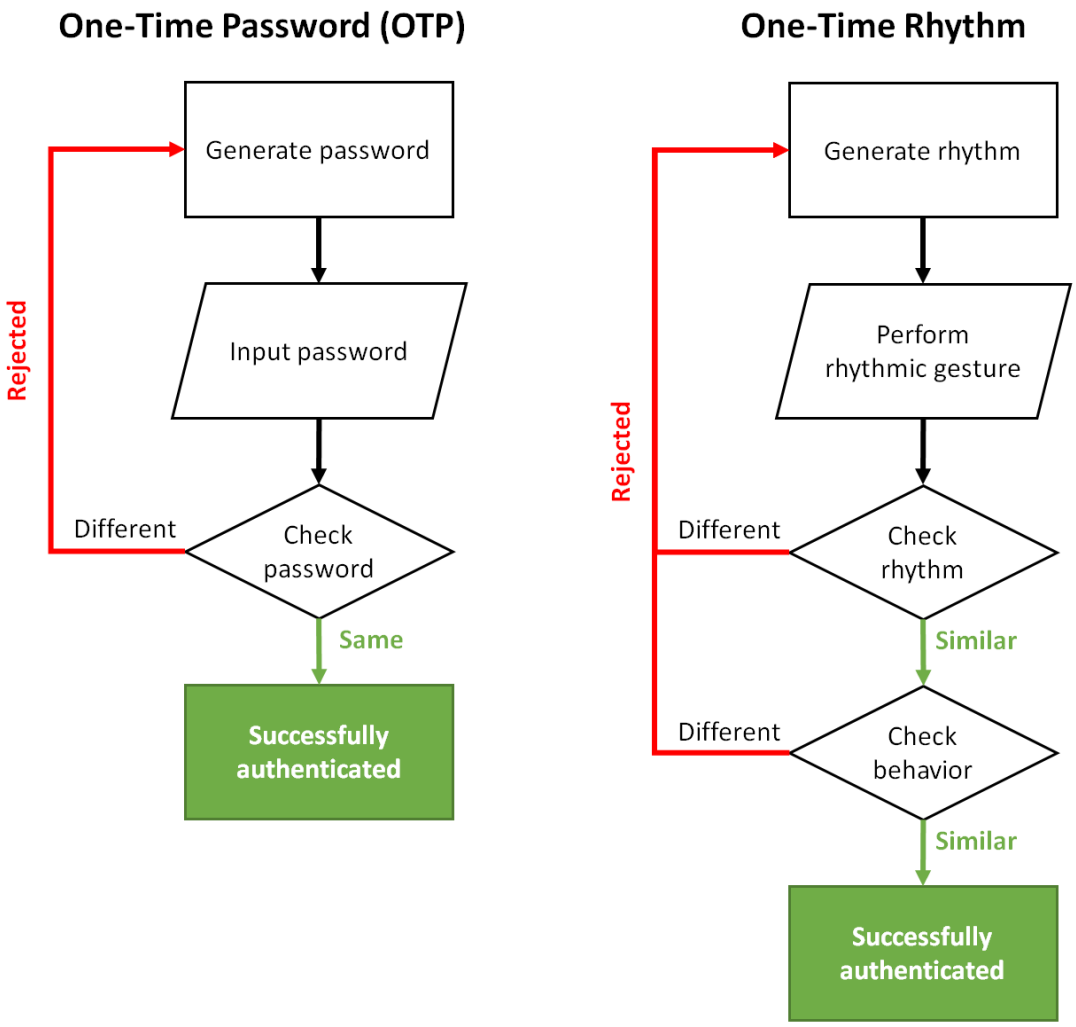


Figure 4.1: Flowchart of OTP and one-time rhythm

4.1.1 External visual and auditory cues

Unlike user-dependent rhythm which the rhythm is produced in the brain, generated rhythm should be produced onto a certain medium so the user is able to shadow the generated rhythm.

Different musical notes are used to test out whether the notes can be suitable to be used for rhythmic gesture. Different tempos have also been implemented to study whether there are any effects in both the rhythm recognition and behavior recognition. It is speculated that the speed and the length of each beat can affect the rhythm recognition as this is dependent on a person's reaction speed and a person's perception of the length of each beat. But these effects also help distinguish the behavior of each and every individual.

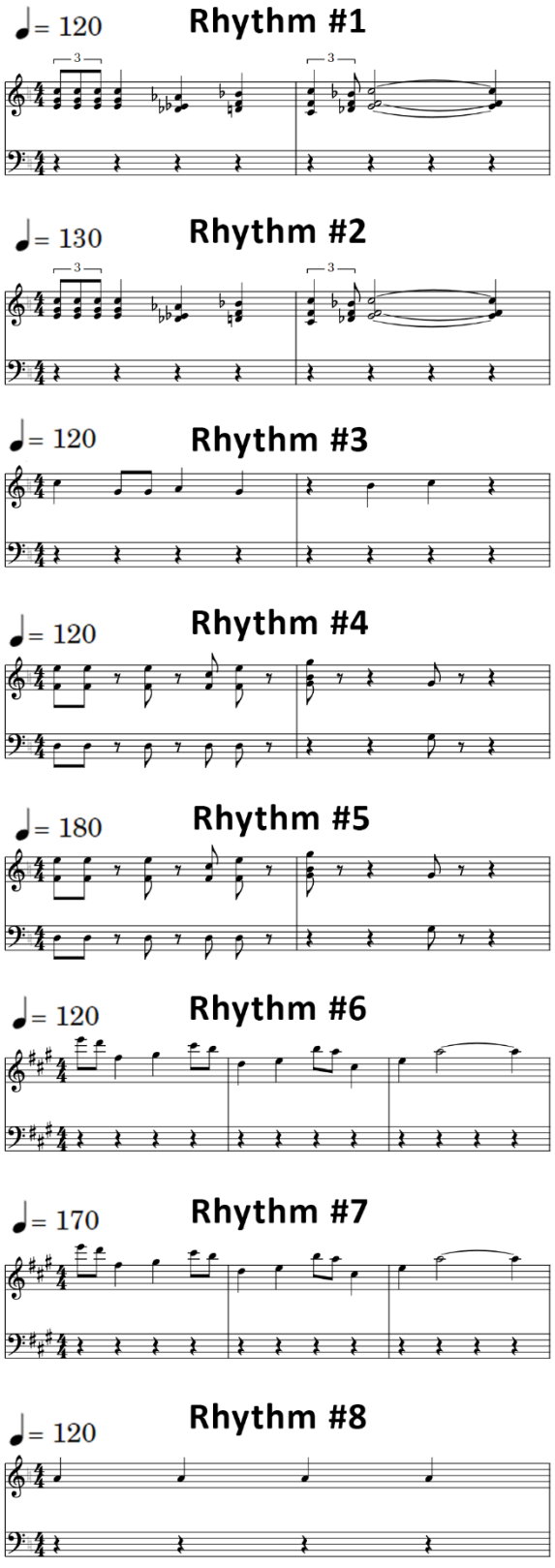


Figure 4.2 Rhythms used for one-time rhythm

Figure 4.2 shows the different 8 rhythms used throughout the experiment for one-time rhythm. All of them are considered different rhythms. Definition of different rhythm is even though some have the same arrangement of notes or the same patterns, they can have different tempos. For example, in Figure 4.2, Rhythm #1 and Rhythm #2 have the same pattern but the beats per minute (BPM) are 120 bpm and 130 bpm respectively; Rhythm #4 and Rhythm #5 have the same pattern but the BPM are 120 bpm and 180 bpm respectively; Rhythm #6 and Rhythm #7 have the same pattern but the BPM are 120 bpm and 170 bpm respectively. The chords shown in Figure 4.2 are the rhythm from their original melodies. These rhythms with the chords were only used to play as auditory cue to the participants during the experiment. However, when the rhythms were used to compare with the rhythmic gesture in the system, they were converted into one single pitch, i.e., all the notes' pitch in each rhythm are change to higher pitch of C6. This is because when using lower pitch, the signal produce from the melody has lower attenuation (slow decay in musical term). Whereas rhythmic gesture using EMG signal has high attenuation after each note or beat. The lower attenuation in signal can cause lower segregation of each note or beat after being applied with integration of signal or Butterworth low-pass filter, causing a high mismatch with the rhythmic gesture of the same rhythm. The higher pitch was not used as auditory cue as it can cause discomfort to some people.

4.1.2 Experimental setups

A total of 7 male participants, ranging between the age of 24 and 30, participated in the experiment. Participants suffering from physical disabilities, pain, and diseases causing potential neural damage, systemic illnesses, language problems, hearing or speech disorders, and mental disorders were excluded. The experiment required all the participants to shadow the same 8 different rhythms. Each rhythm consisted of 5 trials and were randomly performed throughout the experiment. Although all the rhythms were not presented to the participants before the experiment, there were no guarantees that the participants had not known the rhythms beforehand as some rhythms are popular rhythms.

All the rhythms were accompanied with visual and auditory cues. The visual cues are in video form where a screenshot of it is shown in Figure 4.3.

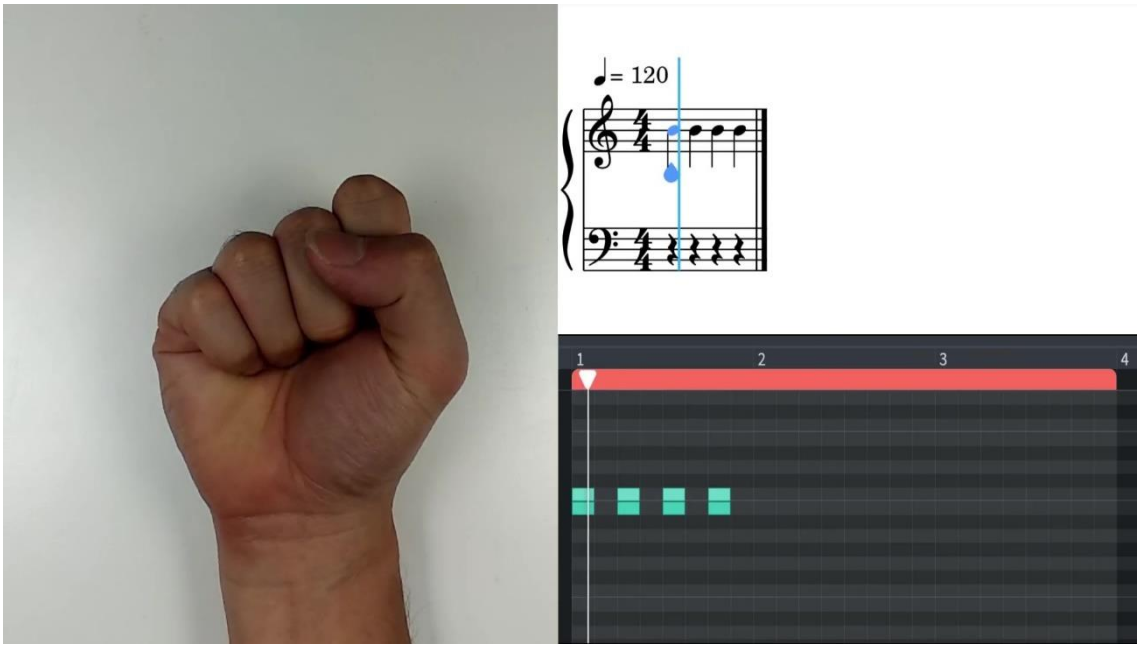


Figure 4.3: Screenshot of the visual cues for one of the rhythms that shows the first movement (left), musical notes (top right), and patterns (top left).

In Figure 4.3, there are three different views of visual cues as different users can have different preferences when trying to shadow and match the rhythmic patterns. The fist view shows how the fist clenches and relaxes with the beat of the rhythm. The vertical line in both musical notes and patterns will move from left to right as the rhythm is being played. While the visual cue is being played, the auditory cue is also played simultaneously matching the rhythm shown in the visual cue.

All participants were asked to shadow the rhythms in front of a display wearing a headphone while seated. All the participants were asked to use only their right hand to perform the rhythmic gesture. Their forearms had all be asked to place on the table and maintained the location for the entirety of the experiment. These measurements were taken to avoid inconsistency in the signal collected from the forearm muscles.

4.2 Rhythm and Behavior

4.2.1 Rhythm recognition

To authenticate using one-time rhythm, the system first needs to match the rhythmic gesture by the user to the generated rhythm, similar to that of OTP which matches the password input by the user to the generated password. Unlike OTP which only matches character by character in the correct order, rhythm requires more than just matching beat.

Rhythm in music consists of beat, tempo, pitch, and amplitude; whereas rhythm produced using EMG consists of beat, tempo, and amplitude, in addition to signal noise.

Since it is impossible to control the pitch through EMG, each generated rhythm will be using only one pitch. Also, amplitude of the rhythmic gesture from EMG can be difficult to match with that of the generated rhythm as it is subjected to how much power the user produced by contracting their muscles while performing each beat. This can be easily affected by how strong or weak an individual is. Therefore, the signal of both generated rhythm and rhythmic gesture from EMG will be converted to binary – 0 for rest or absence of sound, 1 for beat or presence of sound.

To convert both the generated rhythm and rhythmic gesture to binary, the raw signal from both generated rhythm and rhythmic gesture are first filtered using integration of signal. The integrated time for every signal has been set to 32 Hz. This is because the shortest note used in the experiment is eighth note or 8 Hz, and it is split into 4 parts enable the definition of where the note starts and ends. After the integration, the mean of the entire integrated signal will be used as the threshold to define the conversion to binary signal. Hence, any signal above the threshold will be turn into 1 and any signal below the threshold will be turn into 0. The process of converting raw rhythmic gesture signal into binary signal has been illustrated in Figure 4.4. During the conversion to binary signal, the initial point of the signal will start from the first beat of the rhythm. This is to align the signal starting point of both generated rhythm and rhythmic gesture.

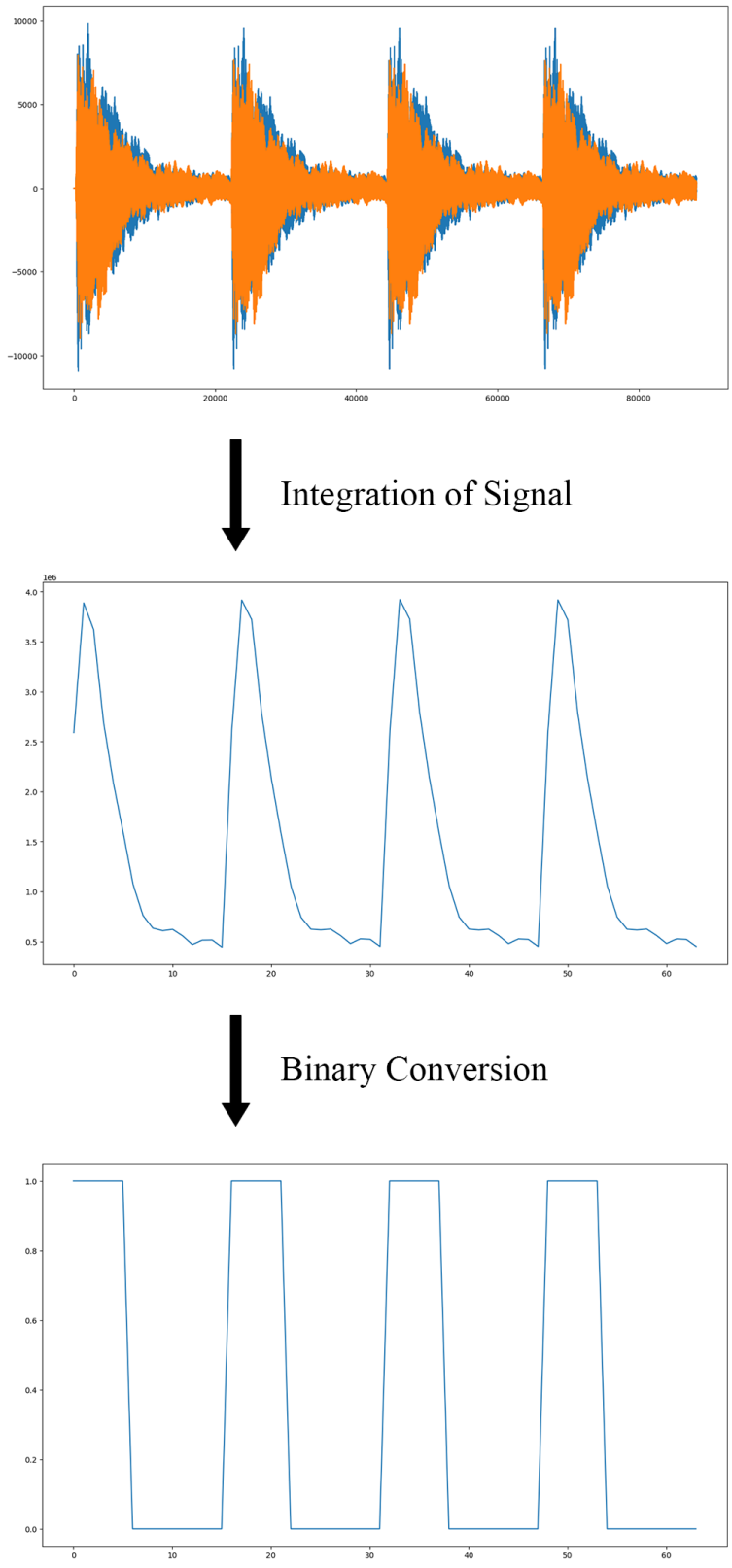


Figure 4.4: Conversion from raw rhythm signal to integration of signal to binary signal.

To find whether one thing is the same or similar to another, similarities or dissimilarities between these two things have to be established to confirm the matches. Levenshtein distance is a technique to find the dissimilarity between two sequences. Levenshtein distance measures the distance or dissimilarity between two sequences using the cost of operation to convert one series to another. There are 3 different operations, delete, insert, and replace, and the algorithm of Levenshtein distance is as follows:

$$Lev(i,j) = \min \begin{cases} 1 + Lev(i-1,j) & delete \\ 1 + Lev(i,j-1) & insert \\ diff(i,j) + Lev(i-1,j-1) & replace \end{cases}$$

An example of Levenshtein distance applied to two different binary signals is shown in Figure 4.5, where the deletion, insertion, or replacement of the signal value has an operation cost of 1. The lower the total cost, the lower the dissimilarity.

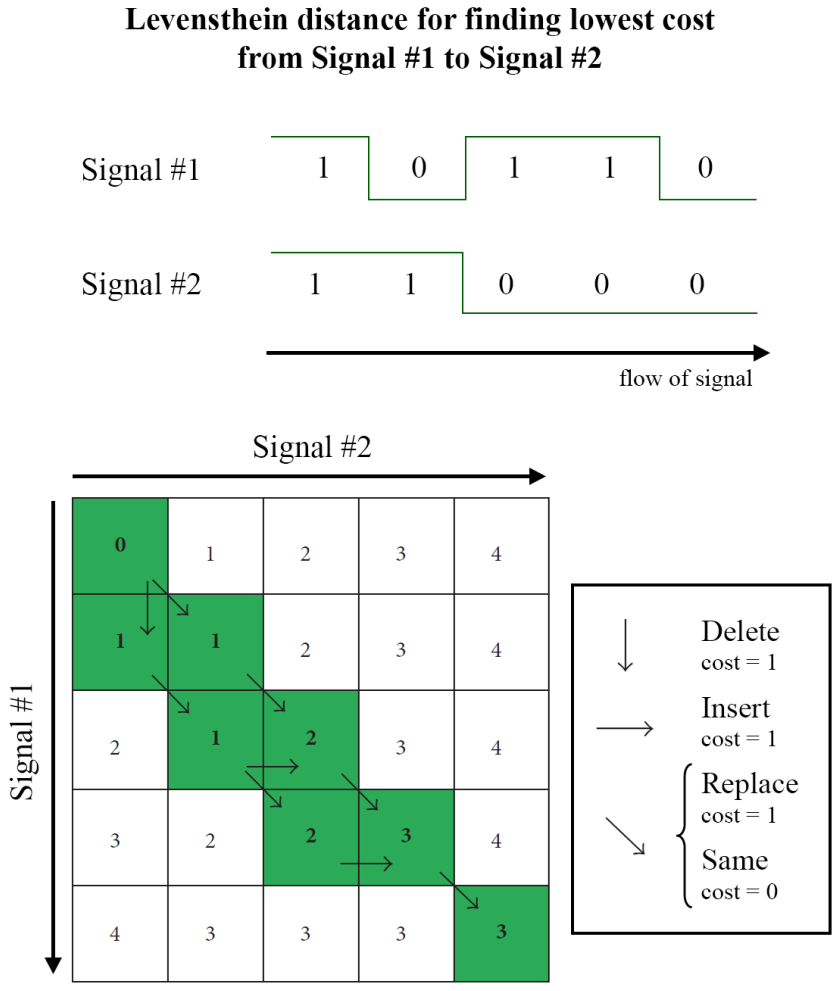


Figure 4.5: Example of Levenshtein distance for finding lowest cost from Signal #1 to Signal #2 where the minimum cost is 3.

In signal processing, a well-known method in determining similarity between two signals is cross-correlation [90]. Cross-correlation measures the similarity of two series by sliding dot product. This means that one signal is sliding through another signal while scanning for the highest level of correlation between them.

When only using one of the methods in analysis, some weaknesses in the measurements have been found in recognizing the binarized rhythmic gesture to the binarized generated rhythm. By using only Levenshtein distance, different rhythms with similar number of peaks at different tempo can produce low cost. This means that by performing different rhythm from the generated rhythm can be considered as the same as the generated rhythm by the system. Whereas using only cross-correlation, if the rhythmic gesture is able to "cover" the area under the graph of the generated rhythm, the algorithm will consider the rhythmic gesture to have maximum similar to the generated rhythm as shown in Figure 4.6, where Signal #2 is able to "cover" all the area under Signal #1 which will output the same value as when comparing Signal #1 with Signal #1.

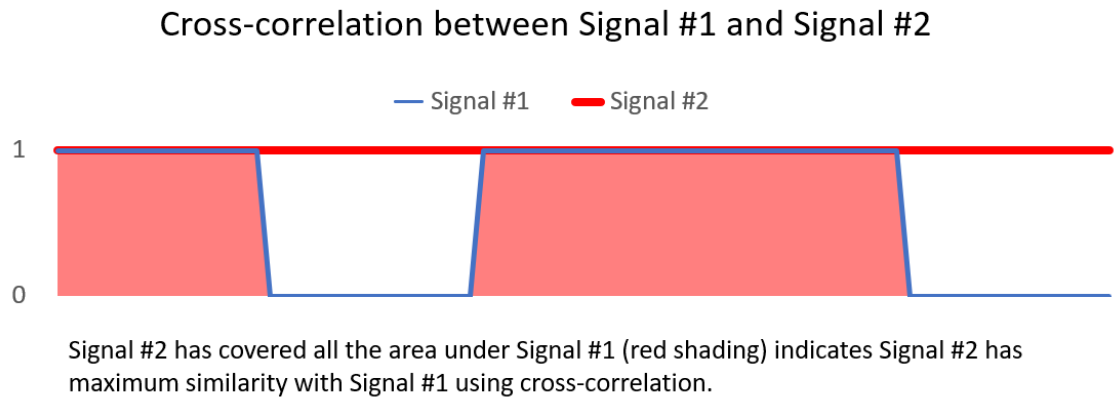


Figure 4.6: Cross-correlation shortcoming where both signals have maximum similarity

Therefore, I have proposed a method which combine both Levenshtein distance and cross-correlation. This can result in a much more optimal measurement in terms of recognizing binary rhythm signal. The equation for combining both Levenshtein distance and cross-correlation is as follows:

$$NS_{Lev} = 1 - \frac{Lev}{N_{gr}}, \qquad if\ Lev < N_{gr}$$

where NS_{Lev} is the normalized similarity of the generated rhythm and rhythmic gesture using Levenshtein distance. lev is the total operation cost output by the Levenshtein distance between two different signals. N_{gr} is the total sample size of the generated rhythm. If the Lev is greater than N_{gr}, then the generated rhythm and the rhythmic gesture can be deemed as completely different. Thus, the rhythmic gesture has to be rejected as impostor.

$$NS_{cc} = \frac{Similarity_{high}}{Similarity_{max}}, \qquad if\ Similarity_{max} > 0$$

where NS_{cc} is the normalized similarity of the generated rhythm and rhythmic gesture using cross-correlation. $Similarity_{high}$ is the highest value output from cross-correlation of two different signals. $Similarity_{max}$ is the maximum value output from cross-correlation of the same signal. If the $Similarity_{max}$ equals to 0, it means that there is no signal from the generated rhythm.

$$Similarity_{pp} = NS_{ld} \times NS_{cc}$$

where $Similarity_{pp}$ is the proposed similarity algorithm. The proposed algorithm uses the product rule of score-level fusion to calculate the estimation similarity of the generated rhythm and rhythmic gestures using Levenshtein distance and cross-correlation, which both methods' similarity outcomes are independent from each other. result should range from 0 to 1, where closer to 0 indicates low similarity, and closer to 1 indicates higher similarity, if and only if NS_{Lev} is greater than 0. If the NS_{Lev} is lower than 0, then it is not required to apply the proposed equation as Levenshtein distance has already confirmed that the rhythmic gesture is totally different from the generated rhythm. Because both methods' similarity outcomes are independent from each other, if one of the methods failed to find any similarity, i.e., normalized similarity of 0, then there should not be any similarity from the proposed algorithm. With this, the proposed algorithm similarity outcome can be seen as more reliable. This is also shown in Figure 4.9 where the EER of the proposed algorithm is 12% which is lower than the EER of Levenshtein distance and cross-correlation, which are 18% and 23% respectively.

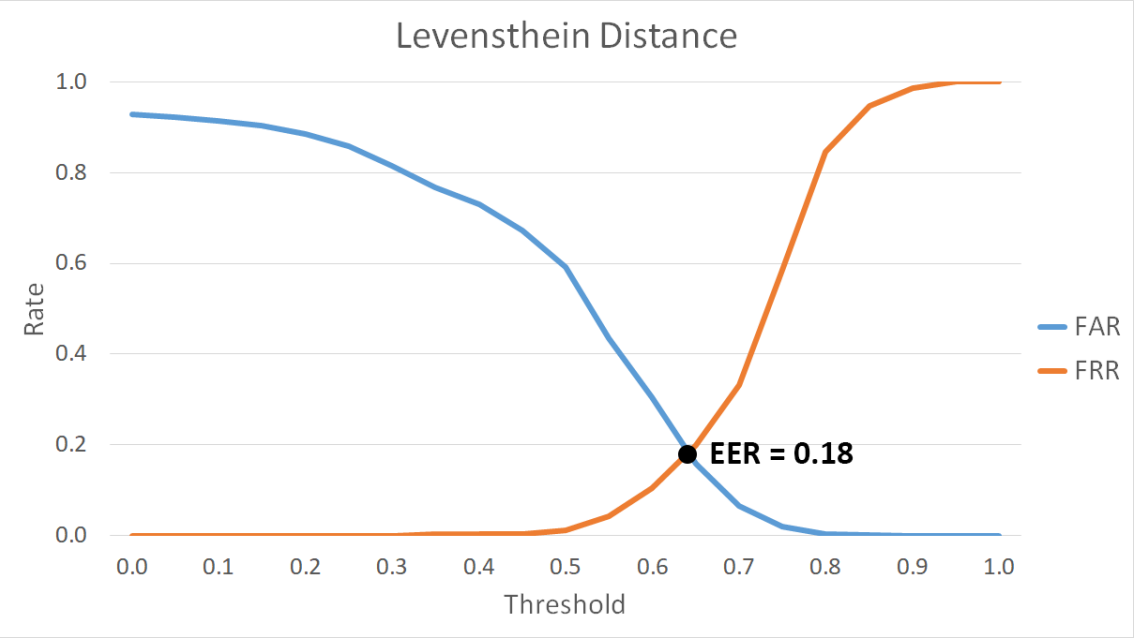


Figure 4.7: FAR, FRR, and EER graph of Levenshtein distance

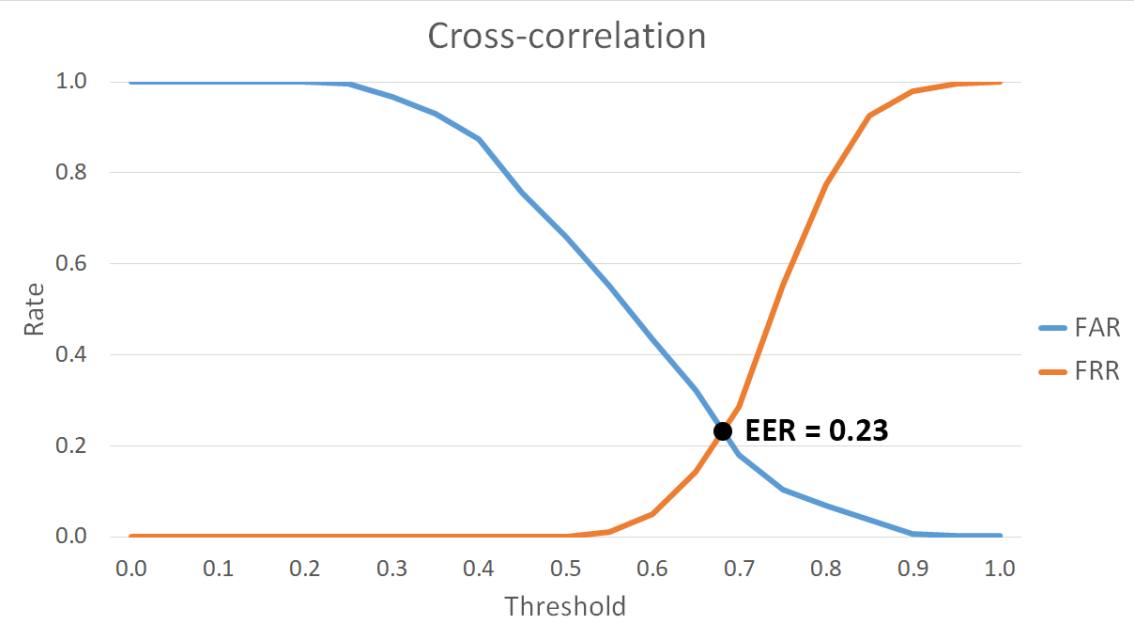


Figure 4.8: FAR, FRR, and EER graph of cross-correlation

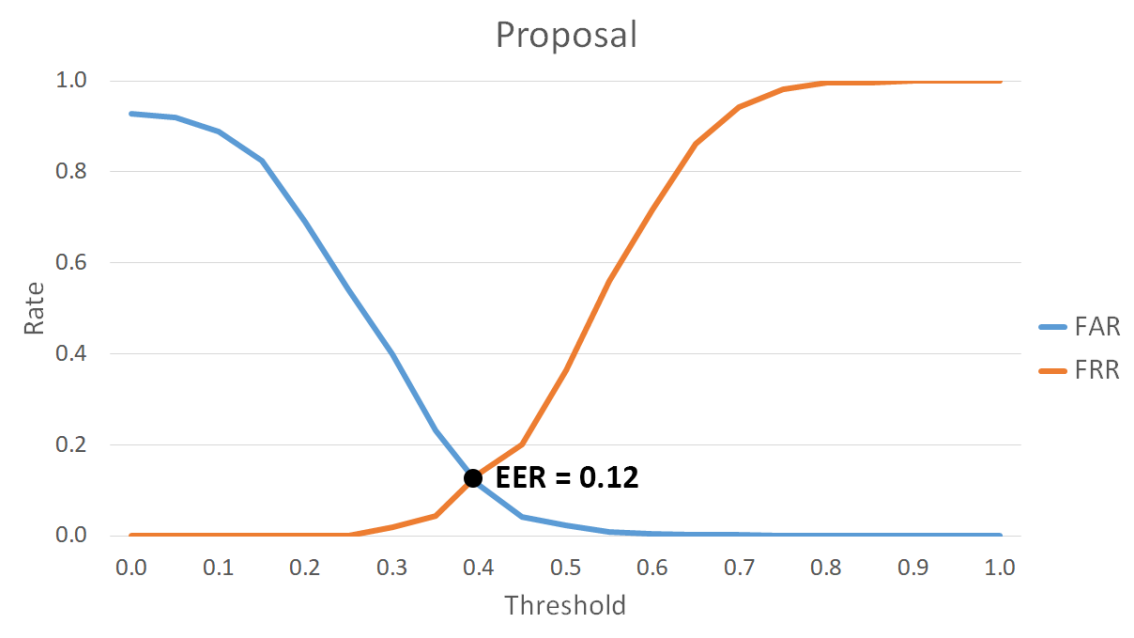
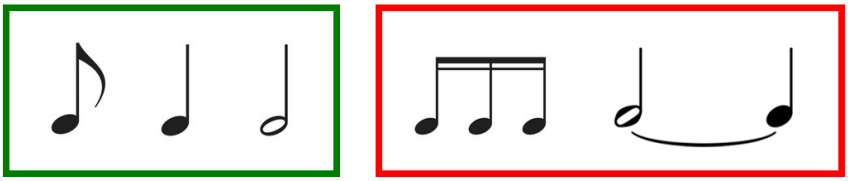


Figure 4.9: FAR, FRR, and EER graph of proposal

Throughout the experiment, it has been discovered that difficult patterns, such as but not limited to short burst beats, and long beats shown in Figure 4.10 (b) can result in lower TAR. These are very dependent on how different users can react to how notes are being played and also the limitation in the sampling rate of the Myo Armband which is only 200 Hz. These notes were included in rhythm #1, #2, #6, and #7 in Figure 4.2. Rhythm #1 and #2 use both aforementioned notes; whereas rhythm #6 and #7

implemented with long beats. The TAR for the rhythm recognition can drop to as low as 37.14% when these notes are implemented to the rhythm using the proposed algorithm.

Thus, it is recommended to use simple pattern such as but not limited to half-note, quarter-note, and eighth-note shown in Figure 4.10 (a). Long rest or pauses are also recommended to be avoid as these notes can cause no behavior in rhythmic gesture since at rest, rhythmic gesture signal stays flat at 0. As for the difference in BPM with same pattern of notes, rhythm #1 and rhythm #2 with 10bpm difference was very difficult to be distinguish by the system. It means that generated rhythm of rhythm #1 has high similarity with rhythmic gesture of rhythm #2, and vice versa. Whereas, rhythm #4 and rhythm #5, and rhythm #6 and rhythm #7, with difference of 60bpm and 50bpm respectively, were able to be distinguished by the system.



(**a**) (**b**)

Figure 4.10: Notes recommended to use (**a**) and notes to avoid (**b**) in generating rhythm

4.2.2 Behavior recognition

The sliding windows technique is used because smaller window size can extract signal waveform feature from the global signal waveform. Window size acts like a high pass filter where subject oriented features can be observed from high frequency region. The subject oriented feature is the behavior of the subject and is hypothesized to be relatively unique. Sliding the window and overlap with the previous window acts as a low frequency feature. This helps the neural network to learn the connection between the windows during the training phase. In short, subject oriented feature is in the high frequency domain which can be observed through the segmentation into smaller window size; whereas the entire rhythm is in the low frequency domain which can be observed through the sliding and overlapping of windows. The system observed the summation of these two domains with the sliding windows technique.

In my research prior to one-time rhythm, it has been confirmed that even when performing the same or similar rhythmic gesture, the system is able to distinguish user based on their behavior through the use of sliding windows technique. In that experiment, all the participants were asked to perform 50 trials of the same rhythm illustration in Figure 4.11. 10-trial data were used for training, whereas the remaining 40-trial data were used for testing. The window sample size was set to 600 and the sliding sample value was set to 1 sample per slide which is illustrated in Figure 4.12. The TAR has achieved at least 91 % which is shown in the confusion matrix of Figure 4.13.

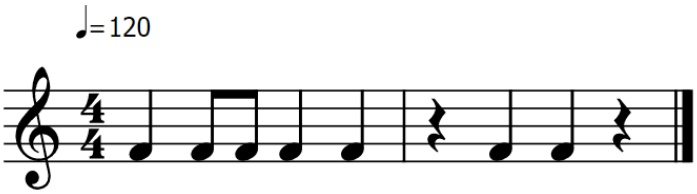


Figure 4.11: Rhythm used in the experiment prior to one-time rhythm research.

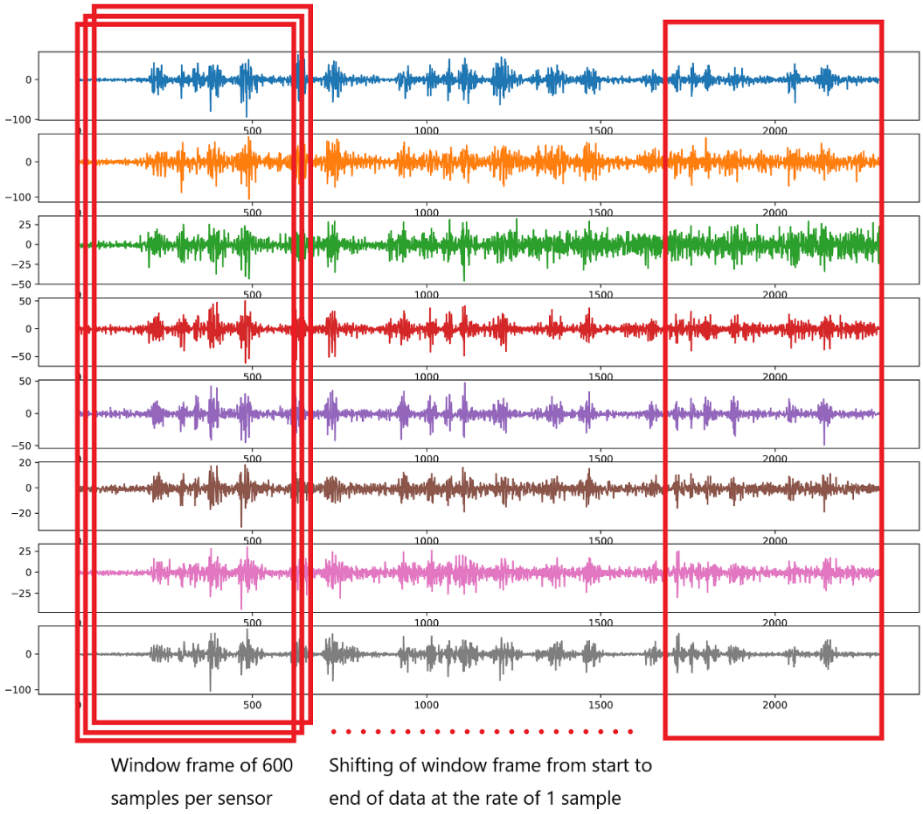


Figure 4.12: Window size and sliding window technique used in rhythmic gesture data.

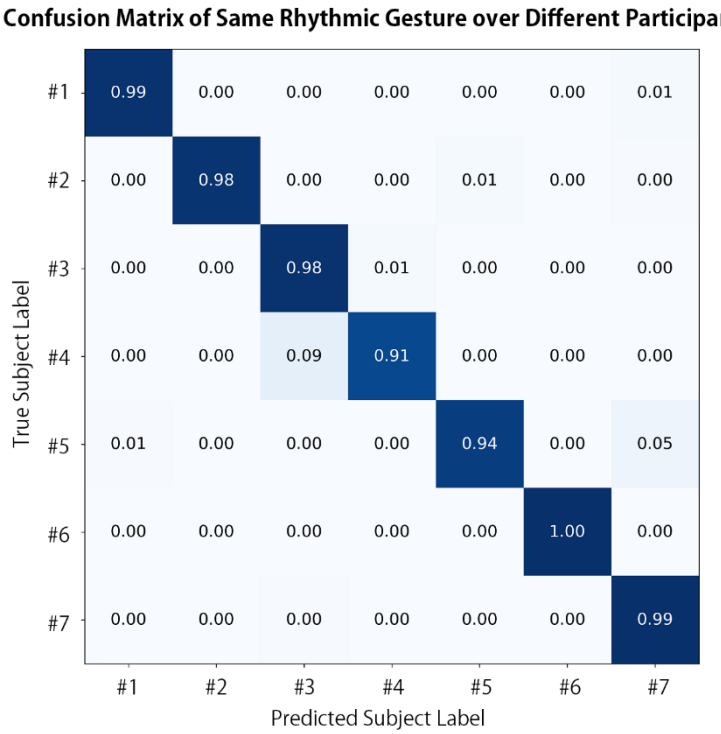


Figure 4.13: Confusion of same rhythmic gesture over 7 different participants.

The rhythmic gesture EMG signals are first filtered using second-order Butterworth low-pass filter. They are then resampled using linear interpolation to 1200 samples before implementing the window method and sliding windows technique.

The window method and sliding window technique is not only used as generating samples of sequences but also acting as a high pass filter that is used for detecting subject-dependent features. This means that the main purpose of using a shorter window length is to exclude full-length features of the rhythmic gesture which exist on the lower frequency band of the signal. In a rhythmic gesture, the frequency contains full-length features of the rhythmic gesture, musical note features, and subject-dependent features, as illustrated in Figure 4.14. The vertical axis of spectral power is applied to the dashed line, where the blue dashed line is the spectral power of full-length rhythmic gesture, the green dashed line is the spectral of musical note, and the orange dashed line is the spectral power of user-dependent features; whereas the vertical axis of gain is applied to the thick semi-transparent line, which represents the filtered gain value, where the blue thick semi-transparent line is the gain of window length of 1200 samples, and orange thick semi-transparent line is the gain of window length of 200 samples. The full-length rhythmic gesture features can be seen to be excluded (red area) when shorter window length of 200 samples is being applied.

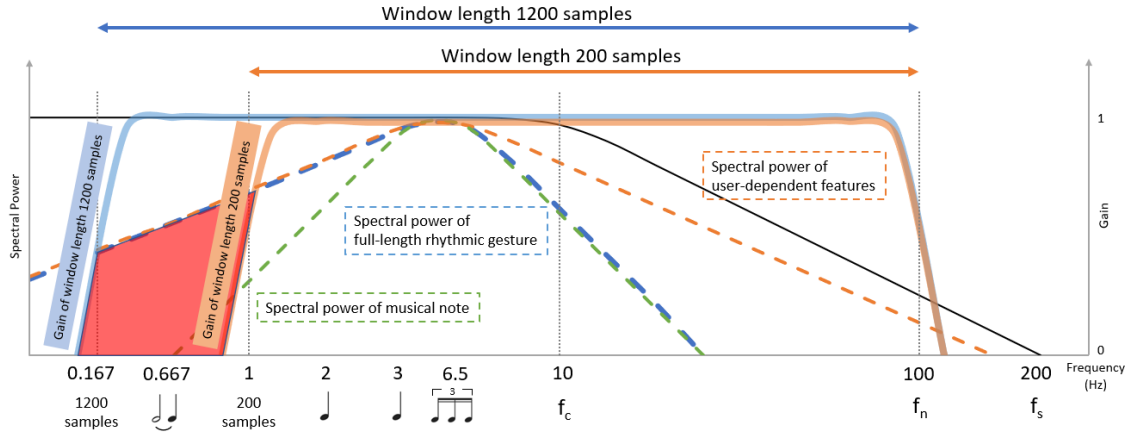


Figure 4.14: Shorter window length excludes the full-length rhythmic gesture features marked with red area.

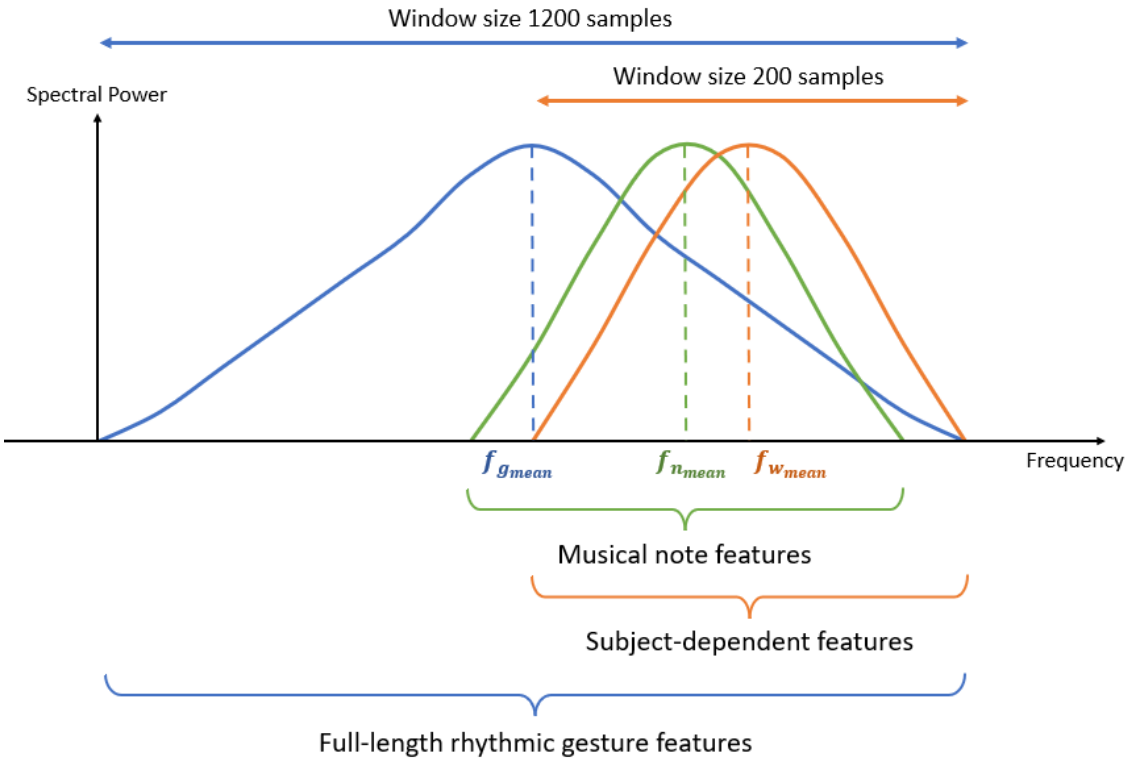


Figure 4.15: The segregation using mean value of full-length rhythmic gesture features, musical note features and subject-dependent features on the rhythmic gesture

When the window size or length of the window used is optimal, it can segregate the higher frequency bandwidth, which contains the subject-dependent features from the lower frequency bandwidth, which is the full length of the rhythmic gesture features as illustrated in Figure 4.15. This can be written with the equation:

$$f_{g_{mean}} < f_{w_{mean}}$$

where $f_{g_{mean}}$ is the mean value frequency of the lower frequency bandwidth, which is the full rhythmic gesture, and $f_{w_{mean}}$ is the mean value frequency of the higher frequency bandwidth which contains subject-dependent features. So, if the mean value frequency of the higher frequency bandwidth that contains the subject-dependent features is larger than the mean value frequency of the lower frequency bandwidth which contains the full-length rhythmic gesture, then the system is able to easily extract the subject-dependent features through the use of the smaller window size. If the window length is too large, the mean value of higher frequency bandwidth and the mean value of lower frequency bandwidth getting closer which increase the difficulty to segregate, making it difficult to extract the subject-dependent features from the rhythmic gesture.

Figure 4.16 shows that when using smaller window size, i.e., 200-sample window, the acceptance rate of authenticate the genuine user – TAR, increase drastically from that of using full-length rhythmic gesture.

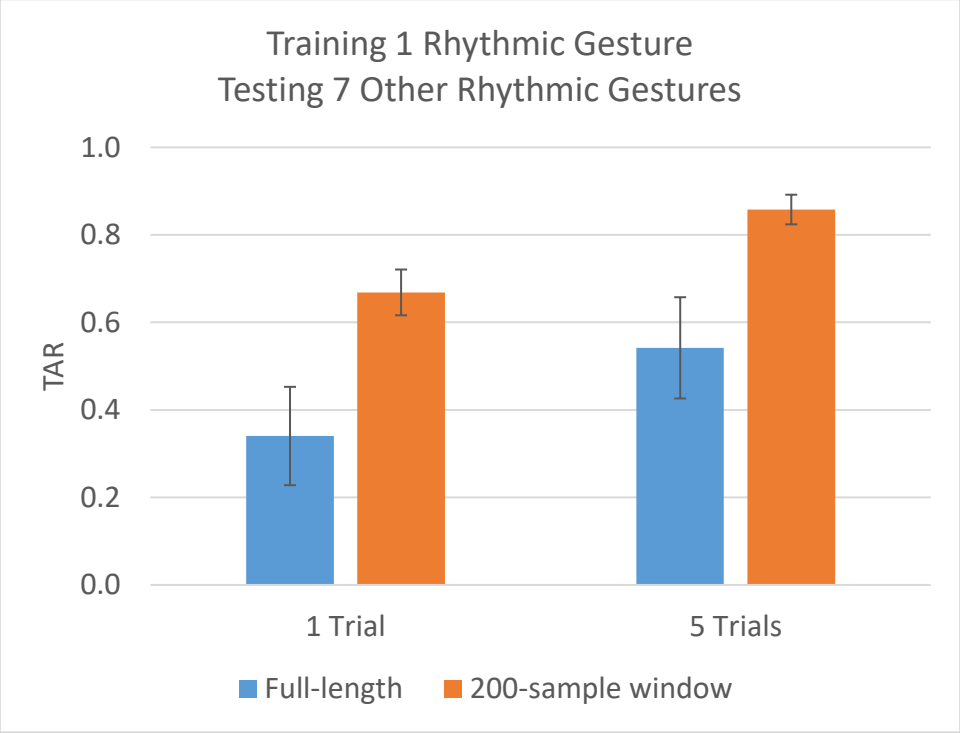


Figure 4.16: TAR of full-length rhythmic gesture and 200-sample window

Sliding window technique is used to observe the change in local spectral of the windowed signal. Sliding window value determines how high or low the frequency bandwidth can be detected. The lower the sliding value, the denser the time domain between two consecutive windows, which delivers a constraint to extract user-dependent features in higher frequency part; whereas, the higher the sliding value means to give constraint sparsely in the time domain between two consecutive windows, which allows

to fit lower frequency part that contains less user-dependent features as compared to full gesture features. This means that lower sliding value of 1 sample will detect higher frequency better than that of higher sliding value of 200 samples with the same window length consequently.

The lowest sliding window value which has a sliding value of 1 can be seen in the equation:

$$E(|f_{(w_i)}|) = E(|f_{(w_{i-1})}|)$$

where $E(|f_{(w_i)}|)$ is the power spectrum of the current window, $E(|f_{(w_{i-1})}|)$ is power spectrum of the window one sliding value forward or "next to" the current window, and E is the operator to obtain the expected power spectrum of the series of $|f_{(w_i)}|$. This equation shows that the sliding window value is denser.

The power spectrum of higher sliding window value which is larger than 1 can be seen in the equation:

$$E(|f_{(w_i)}|) = E(|f_{(w_{i-\partial})}|), \qquad \partial > 1$$

where $E(|f_{(w_i)}|)$ is the power spectrum and phase spectrum of the current window, $E(|f_{(w_{i-\partial})}|)$ is the power spectrum and phase spectrum of the window ∂ sliding value forward the current window, ∂ is the term of cyclic signal, and E is the operator to obtain the expected power spectrum of the series of $|f_{(w_i)}|$. This equation shows that the sliding window value is sparser than the sliding window value of 1.

The sliding values are tested at a shifting rate of 1200 samples with window size of 1200 samples (full-length rhythm), and 200 samples, 100 samples, 50 samples, 20 samples and 1 sample with window size of 200 samples (window method), listed out in Table 4.1. The sliding windows of 200 samples has no overlapping; whereas 100, 50, 20, and 1 sample(s) will produce overlapping of the windows. This overlapping helps to prevent discontinuity between two consecutive windows. This means that the conditionality of the sliding window method with no overlap interval is that time-series causality between different windows is not or is unable to be learned, i.e., time-series causality between different windows does not contribute to the distance in the signal space. Figure 4.17 illustrates the sliding window technique and its terminology.

As illustrated in Figure 4.18 and Figure 4.19, smaller window size paired with smaller sliding value and overlapping significantly increases the TAR of user recognition through their behavior. As sliding window shift at a lower rate, the TAR also increases. This is because the lower the rate of shifting, the higher the subject-dependent features can be observed. These subject oriented features are the unique behavior that the user projects. When there is no overlapping during the sliding window, the TAR is lower. This also helps to confirm that overlapping of windows helps observe the continuity of the behavior better. Therefore, the analysis of the behavior recognition has been conducted with window size of 200, sliding value of 1 with overlapping.

Table 4.1: List of window size, sliding value of window, and overlapping of window being tested.

X-axis	Window size (sample)	Sliding value (sample)	Overlapping
Full gesture	1200	1200	No
200	200	200	No
100	200	100	Yes
50	200	50	Yes
20	200	20	Yes
1	200	1	Yes

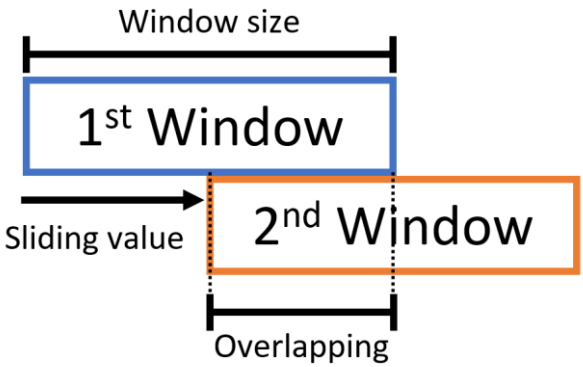


Figure 4.17: Window size, sliding value, and overlapping of window

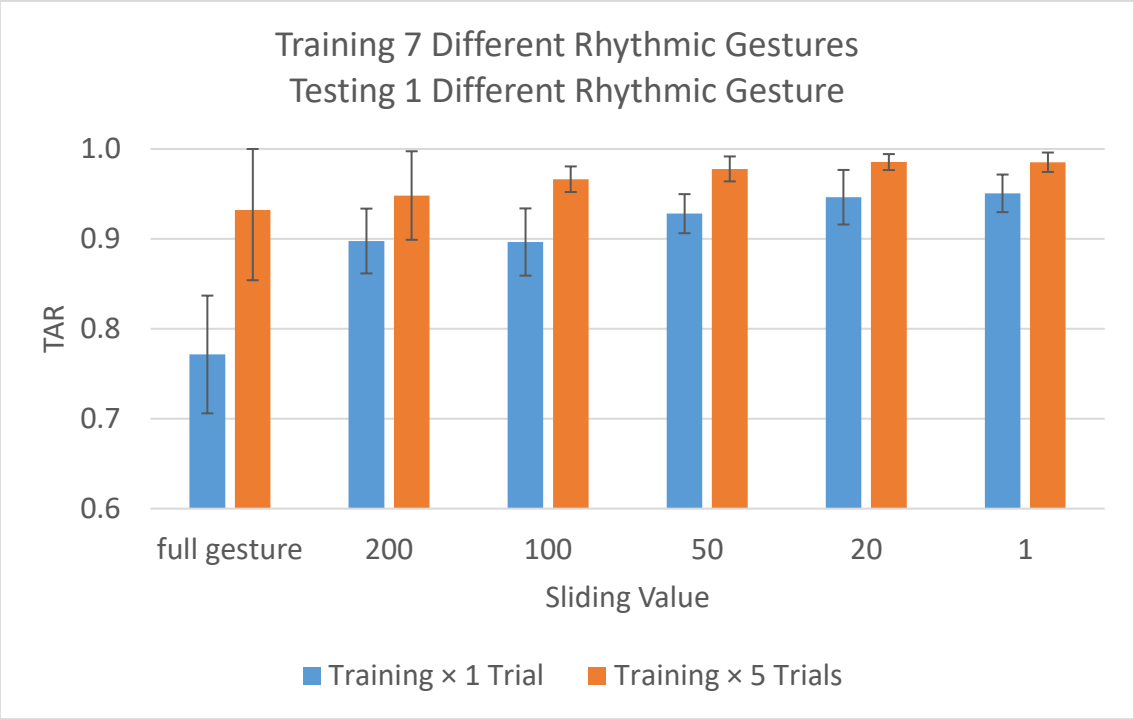


Figure 4.18: TAR of behavior recognition through different window size and overlapping, where training data consist of 7 rhythm gestures and testing consist of 1 rhythm gesture not available in the training data.

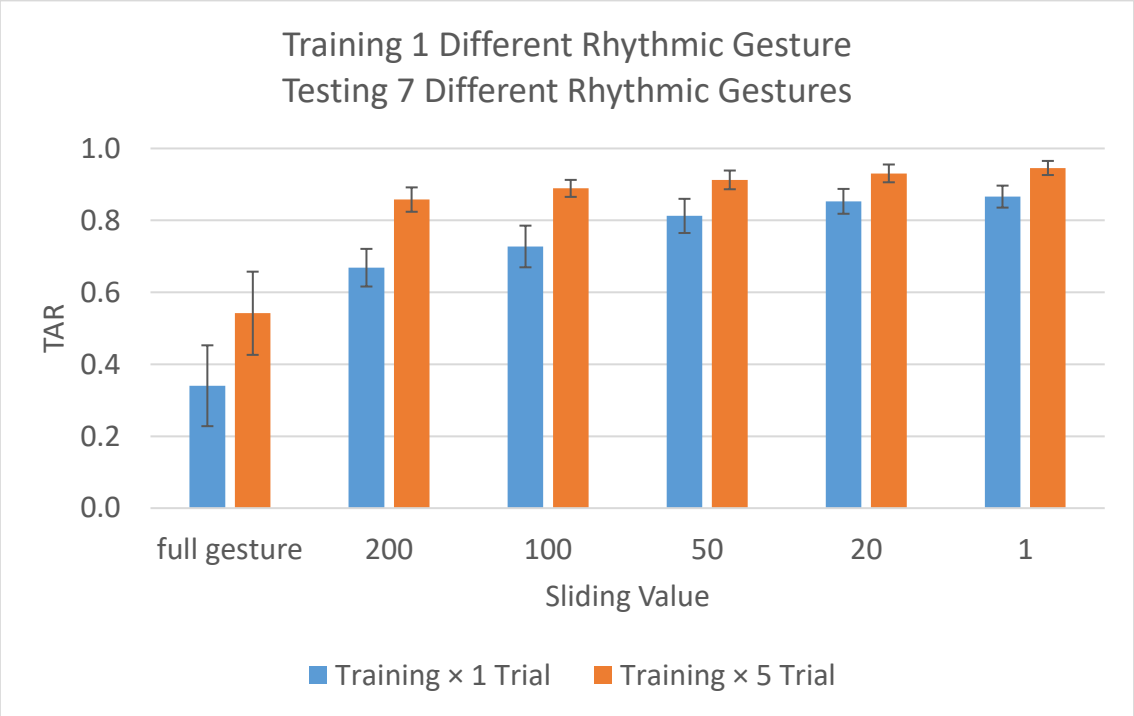


Figure 4.19: TAR of behavior recognition through different window size and overlapping, where training data consist of 1 rhythm gesture and testing consist of 7 other rhythm gestures not available in the training data.

As aforementioned, from the previous experiment, it has been established that even when using the same rhythmic gestures, the system was still able to distinguish users through their behavior. But how well can different rhythmic gestures perform – in terms of rhythm used for testing is totally different from rhythm used for training, has not been discussed. A comparison of same rhythm against different rhythms or as being known in this dissertation – one-time rhythm has been evaluated as shown in Figure 4.20. Note that the same rhythm data in the current experiment is a new set of data different from that of the previous experiment.

From the experiment setup, there were a total of 8 different rhythms used. Each rhythm contains 5 trials. In the same rhythm evaluation, all 8 different rhythms and 4 of the trials in each rhythm were used as training data; whereas the remaining 1 trial was used as testing data. Cross-validation has been implemented resulting in 40 sample results produced.

As for one-time rhythm, 4 of the total 8 different rhythms with 4 of the trials in each of the 4 rhythms were used as training data; whereas the 1 trial from one of the remaining 4 rhythms was used as testing data. Combination algorithm has been used to rearrange the training data and testing data throughout the evaluations, resulting in a total of 1400 sample results.

As for data processing, neural network from the TensorFlow Keras Sequential model was used to predict the result. The setup of the neural network is 8 inputs, which are the 8 different channels of the EMG sensors of each gesture, 10 hidden layers with 100 nodes for each layer, and 20 epochs. The loss function used is categorical cross entropy, the optimization function used is ADAM, and the batch size used in the training phase is 1000.

The similarity or dissimilarity of the features extracted from the window method is being computed using neural network. The neural network used is similar to that of the user-dependent rhythm which output the estimation value or probability on all the label. The rhythmic gesture is first split into smaller window chunks using the window method and the sliding window technique. The total number of window chunks for each rhythmic gesture can be obtained using the equation:

$$WC = \frac{(SS_{RG} - SS_{WS})}{SV} + 1$$

where WC is the total window chunks, SS_{RG} is the sample size of one rhythmic gesture, SS_{WS} is the sample size of window size, and SV is the sliding value of sliding window.

Labels are given to each window chunks, i.e., from 1 to 7 as there are a total of 7 participants, before the training and testing phase, and after the testing phase, an estimation will be output by the neural network ranging from 0 to 1. The highest estimation value will be selected as having the highest similarity. One difference from that of user-dependent rhythm is the estimation output does not reflect the entirety of a rhythmic gesture, but instead the window chunks of the rhythmic gesture and are

independent from each other. The estimations of all the window chunks of a rhythmic gesture are then added up to predict which participant does the rhythmic gesture belongs to by using the "winner-take-all" method. For example, a rhythmic gesture is divided into 20 window chunks; after testing, the prediction output 15 of the 20 window chunks to be label 1, i.e., participant #1, and by using the "winner-take-all" method, the resulting rhythmic gesture will be predicted as participant #1's rhythmic gesture. And if the label in the tested rhythmic gesture is labeled as 1, then it has been correctly matched, else if the label is not 1, then it is wrongly matched.

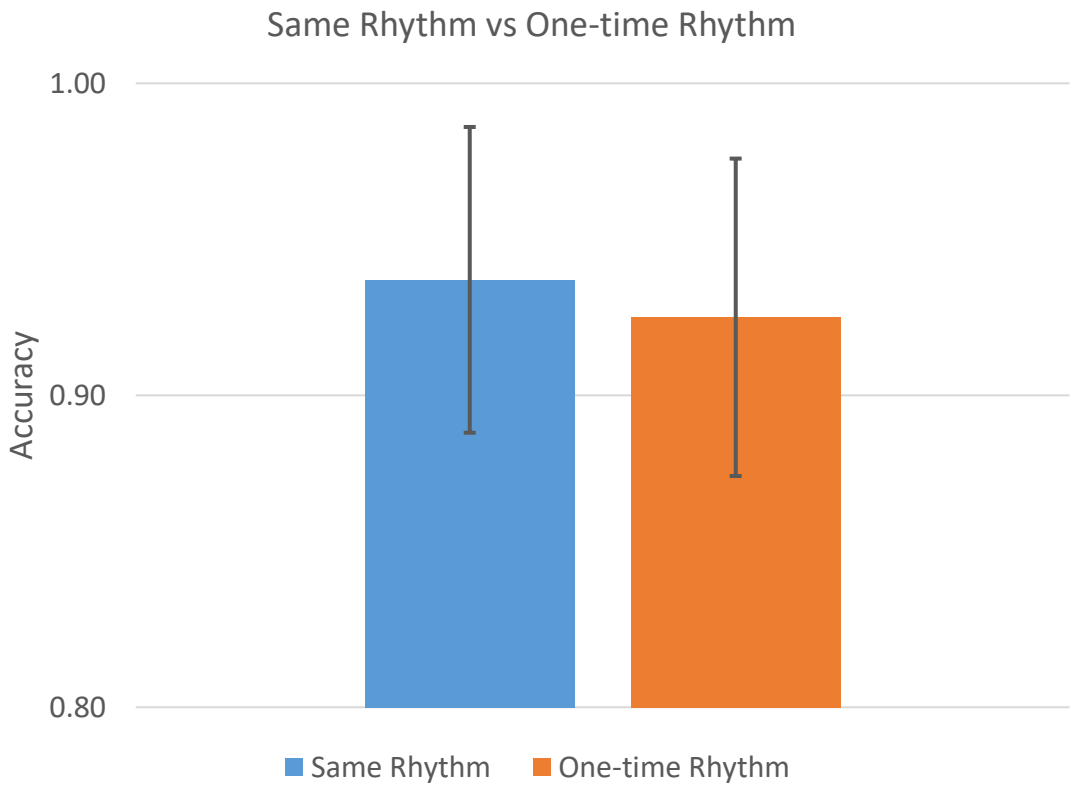


Figure 4.20 Comparison of average TAR and deviation of same rhythm and one-time rhythm.

Even though the TAR and its deviation seem similar, it can still be argued that the sample from each result can vary significantly. Thus, ANOVA has also been used to evaluate whether the samples within each result differs significantly.

A one-way ANOVA has been performed and shown in Table 4.2(b) to compare the effect of same rhythmic gesture and one-time rhythmic gestures on TAR. The one-way ANOVA revealed that there is no significant difference in mean TAR between same rhythm and one-time rhythm ($F(1, 1438) = [1.8633]$, $p = 0.1725$):

- *P-value* > alpha, *0.1725* > 0.05
- *F-value* < F-critical, *1.8633* < 3.8479

Table 4.2 (a): Summary data for one-way ANOVA

Summary

Groups	Samples	Sum	Average	Variance
Same Rhythm	40	37.4797	0.9370	0.0027
One-time Rhythm	1400	1294.9586	0.9250	0.0030

Table 4.2 (b): One-way ANOVA on same rhythm and one-time rhythm

One-way ANOVA

Source of Variation	Sum of Squares	df	Mean Square	F	P-value	F-crit
Between Groups	0.0056	1	0.0056	**1.8633**	**0.1725**	3.8479
Within Groups	4.3379	1438	0.0030			
Total	4.3435	1439				

Lastly, a threshold should be applied to the behavior recognition to either accept the rhythmic gesture being performed by the genuine user or reject the rhythmic gesture as being performed by impostor.

4.2.3 Rhythm Recognition × Behavior Recognition

Combining both rhythm recognition with behavior can help strengthen the security of the system. To first recognize rhythm, the user will have to own certain device to be able to generate a rhythm under a unique token. This can be considered as ownership factor. If the device is stolen or the token is compromised, impostor will be able to generate a rhythm using that unique token, an authenticate as the genuine user. But by implementing the behavior recognition, even if rhythm recognition is accepted by the system, since the generated rhythm matches the rhythmic gesture, the behavior of who performed that rhythm can be very different, resulting in impostors unable to be authenticated by the system. This behavior is the inherence factor. Therefore, rhythm recognition and behavior recognition are a multi-factor authentication without extra actions from the user as compare that to convention OTP.

The overall performance metric of one-time rhythm is a score-level fusion using the product rule on the performance metric of proposed similarity in rhythm recognition and the performance metric of rhythmic gesture behavior from behavior recognition. The equation is as followed:

$$PM_{oa} = PM_{ps} \times PM_{bh}$$

where PM_{oa} is the overall performance metric, PM_{ps} is the performance metric of proposed similarity, and PM_{bh} is the performance metric of the rhythmic gesture behavior. The performance metric is TAR, FAR, and EER. The overall performance of one-time rhythm uses product rule because the performance of rhythm recognition outcome and behavior recognition outcome are independent from each other.

Since one-time rhythm is proposed to reduce the chance of being accessed by impostor, the overall FAR is improved through the equation. For example, FAR of rhythm recognition is 0.1 and FAR of behavior recognition is 0.1, then the overall FAR of one-time rhythm is reduced to only 0.01. But this also brings drawback to the overall TAR. For example, TAR of rhythm recognition is 0.9 and TAR of behavior recognition is 0.9, the overall TAR of one-time rhythm will be 0.81. This means that one-time rhythm improves the security by reducing the chance of impostor from authenticate as genuine user but at the cost of lowering the chance of genuine user being accepted. This use case is highly important in public domain such as workspace, public devices, confidential data, and others.

4.2.4 No learning curve required for users

Since one-time rhythm generates a new rhythm every time, the rhythms can be unfamiliarized to the users. This unfamiliarity usually affects the performance of the user in what they are doing. Usually when introduced with a new technique or technology, there could be a requirement for the users to practice and increase the experience of using the technique or technology in order to increase the performance or the proficiency of using the them. For example, when first typing on a keyboard, user can have low performance in typing but with some practice, the performance can be improved. This relation between the performance and the amount of practice is known as learning curve.

In rhythmic gesture, the definition of learning curve is with repetitive practice of the same rhythm, users should have learned the rhythm sufficiently to improve their performance in performing in that particular rhythm. In the one-time rhythm experiment, each rhythm has been performed and recorded five times from each participant. This means that on the fifth trial of performing the same rhythm, the performance, i.e., TAR should have increased significantly as compared to the TAR on the second trial. For example, the TAR of Rhythm #1 in the fifth trial should be higher than that of Rhythm #1 in the second trial. This is because in the fifth trial, the users have learned Rhythm #1 five times, whereas during the second trial, the users have only learned Rhythm #1 for the second time.

In Figure 4.21, T# indicates the trial number, thus T1 indicates trial number 1; whereas T1-T2 indicates trial number 1 being used as enrollment, and trial number 2 being used as probe. This means that in T1-T2, user has only learned the rhythm for the second time; whereas in T4-T5, user has already learned the rhythm for the fifth time. Supposedly, as trials proceed, the users should be more proficient to the rhythm and

should have produced higher performance. However, the trend of TAR shown in Figure 4.21 does not change significantly as the trials proceed; in fact, when trial number 1 was used as training data, trial number 2 as testing data, produced the highest TAR as compared to later trials. This means that even without repetitive practice, which is the case in one-time rhythm, it should not be detrimental to the TAR of the biometric system.

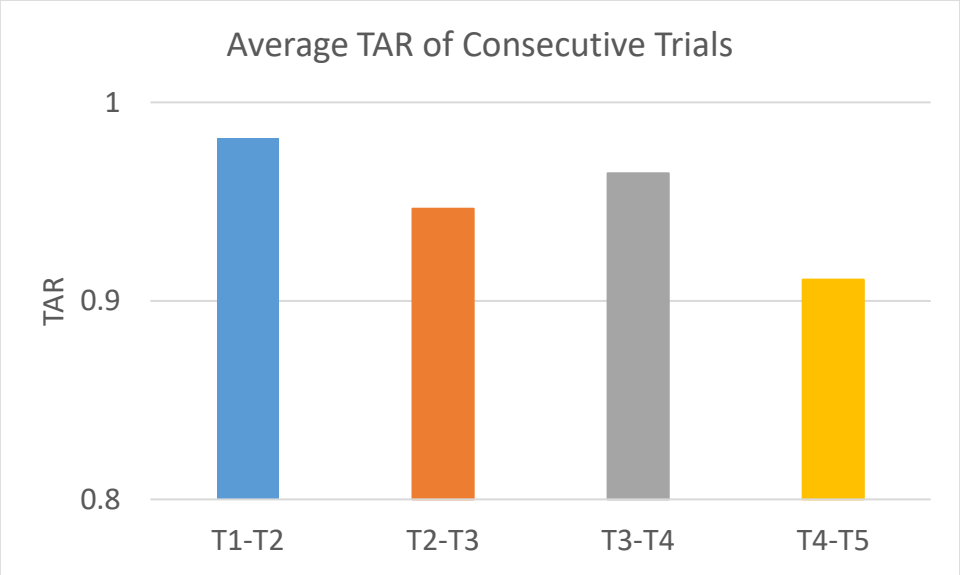


Figure 4.21: Average TAR of consecutive trials

4.3 User-dependent rhythm versus one-time rhythm

From Chapter 3, each user creates their own rhythm, which is known as self-dependent rhythm; whereas in Chapter 4, a distinct rhythm is being generated by the system every time at the beginning of the authentication process, one-time rhythm. Both techniques have their advantages and disadvantages.

User-dependent rhythm:

- Advantages:
    - Convenient – from user memory, does not require extra generator.
    - Less computation power – Uses full gesture to authentication, no smaller size window and sliding window technique required.
- Disadvantages
    - Unable to fully utilized subject-dependent features – Difficult to implement sliding window due to chance of long pauses in rhythm created by the user. These pauses from the signal point of view are usually flat lines close to 0 which contains no subject-dependent features.

    - Less secure – Chance of being stolen and practiced since the rhythm used is always the same.

One-time rhythm:

- Advantages:
    - Able to use different rhythms
    - Able to utilized subject-dependent features – sliding window technique can be used to detect subject-dependent features as generated rhythm can be programmed to remove inconsistency and unwanted long pauses in the rhythm.
    - Lower chance of being stolen or mimic – the rhythm is always different
- Disadvantages of one-time rhythm:
    - Less convenient – requires generator and an additional step for the user to perform in the authentication process
    - Higher computational power – it uses smaller window size and sliding window technique

Based on the aforementioned advantages and disadvantages, it can be concluded that user-dependent rhythm can be used in private scenarios, such as personal devices or home; whereas one-time rhythm can be used in a less private scenarios such as building entrance that can require higher security measures. Of course, one-time rhythm can also be used in the same scenario as user-dependent rhythm, as long as the user does not mind the hassle of performing extra steps in the authentication process that is generating and shadowing the rhythm.

4.4 Summary

One-time rhythm has been proposed to increase difficulty in stealing or mimic rhythmic-based dynamic hand gesture. One-time rhythm generates different rhythms every time when a user starts the authentication process. One-time rhythm uses both rhythm recognition and behavior recognition to complete the authentication process. Because the rhythm is always different, similar to that of OTP where the password is dynamic, the rhythm recognition is considered as a dynamic too. And as mentioned, behavior is always different thus making it also dynamic. Consequently, one-time rhythm is a dynamic × dynamic authentication, which in theory difficult to be stolen or mimic as two methods are not always the same.

For the rhythm to be recognized, both the generated rhythm and the rhythmic gesture will first be converted to binary signal. Through binary signal, a proposed similarity algorithm which combines both Levenshtein distance and cross-correlation has

been applied to find the similarity between the generated rhythm and rhythmic gestured. If the results from the proposed algorithm is over the threshold set in place, the rhythmic gesture will be deemed similar to the generated rhythm, and vice versa.

After the rhythmic gesture is able to match with the generated rhythm, the system will then proceed to evaluate the behavior of the rhythmic gesture. Smaller window size and sliding window technique has been implemented to detect subject dependent feature. Higher value in result will likely deem the rhythmic gesture is being performed by the genuine user; in contrast, lower value in result will likely deem the rhythmic gesture as impostor. The overall performance of one-time rhythm is the score-level fusion using product rule on rhythm recognition and behavior recognition. From the equation, it can be seen that the FAR greatly reduces with the drawback of reducing the TAR of genuine user being accepted too. This also indicates that one-time rhythm can be more suitable to be use in public domain or devices.

Lastly, it is also demonstrated that there is no learning curve for user to shadow a generated rhythm. The TAR does not depend on how many times the user has performed the rhythmic gesture. This means that even when new rhythm is generated and the user has never done the rhythm before, they will still be able to perform the rhythmic gesture with acceptable result.

Chapter 5

Conclusion

Rhythmic-based dynamic hand gesture with electromyography (EMG) has proven to be a robust in terms of reliability, and security. When comparing with other behavioral biometrics, rhythmic-based dynamic hand gesture with EMG has less external or outside factors affecting the performance of the biometric modality due to the direct contact of the EMG sensors to the skin. This means that there is a lower chance of failure-to-acquire (FTA) and failure-to-enroll (FTE) too. This direct attachment also increases the difficulty for a third person to perform measurement. Rhythmic gesture does not require space to perform resulting in the ability to be hidden while performing, in addition to perform even when the hands are busy. Based on the aforementioned, it proves that rhythmic-based dynamic hand gesture with EMG can be more reliable and secure than other biometrics qualitatively.

As for quantitatively, reliability and security can also be seen from self-dependent rhythm with EER of as low as 1.32 % and is considered low when compared with other behavioral biometrics shown in Table 3.2. Physiological biometrics is popular in terms of authentication due to its convenience but it is not without any fault. This is especially when it comes to being compromised or stolen. As mentioned, physiological biometrics is hardly changeable. Rhythmic-based dynamic hand gesture depends on the behavior of the user, therefore even if the rhythm is known or stolen, impostor is still unable to be authenticated as genuine user due to the difference in how the rhythmic gesture is being performed. This has been shown when one of the experiments simulated rhythm being compromised or stolen, rhythmic-based dynamic hand gesture has low FAR of (10.38 ± 1.79) % as compared to compromised physiological biometrics such as fingerprint which has FAR of over 60 % [1].

One-time rhythm has also been proposed, where different rhythm is being generated by the system every time for the authentication process. One-time rhythm is split into rhythm recognition and behavior recognition. A proposed method of combining Levenshtein distance and cross-correlation through product rule of score-level fusion has been proposed and the result of matching rhythmic gesture to generated rhythm is more reliable than of Levenshtein distance or cross-correlation only. This can be seen in the EER of the proposed similarity measurement which is at 12 %. As for behavior recognition, smaller window size and sliding window technique have been used to detect subject-dependent features and has been proven to be able to pick up these features even when different rhythmic gestures are use. This has been proven when comparing same-rhythm and different-rhythm, where the TAR and variance of both have insignificant differences. The overall performance of one-time rhythm also uses the product rule of score-level fusion. This is because rhythm recognition and behavior recognition are independent from each other. This resulted in a higher security as it can help reduce the FAR but at the cost of TAR. But based on the use case, this can be more favorable when

being used in public domain where less chance of being stolen or mimicked is more important.

Both user-dependent rhythm and one-time rhythm have their own advantages and disadvantages. Because user-dependent rhythm is more convenient and require less computational power, it can be suitable for personal device; whereas one-time rhythm can reduce the chance of being stolen or mimicked, it can be suitable for less private domain such as building entrance.

Rhythmic-based dynamic hand gesture with EMG has been proven to be a viable biometric authentication method with its versatility, ease of use, and security. Moreover, different authentication method can also be fused with rhythmic-based dynamic hand gesture, such as the case of one-time rhythm.

5.1 Limitation and future work

There are some limitations in the of the current rhythmic-based dynamic hand gesture, such as the approach used in the neural network is identification-oriented, where rhythmic-based dynamic hand gesture is much suitable for verification-oriented. For the system to be realized as a verification-oriented neural network, one important measurement is that the system should be able to determine whether the target subject is who he/she claims to be. Which also means that the system should be able to reject impostor. To do so, the target subject (acting as an impostor or an unknown individual), has to be omitted from the training set and should not be included in the enrollment (as he/she is an unknown individual not enrolled yet). Since verification is the process of verifying an individual identity, or one-to-one matching, the target subject will have to verify himself/herself as an individual who is available in the enrollment. This should end with a rejection from the system as the target subject is an impostor and not the person he/she said to be.

Based on my current research, what is called “training data” is enrollment. The enrollment in my research has also been used as training set which has been used to train the classifier of the deep neural network. Whereas, what is call “testing data” is a probe which is usually used as the test set. Enrollment and probe should not contain the same data. Based on the arrangement of the structure of enrollment, training set, and probe (test set) in my research, it is an identification-oriented neural network which act as a validation for internal procedure to achieve verification.

For the system to work as a verification-oriented neural network, the target subject (acting as an impostor or an unknown individual), has to be omitted from the training set and should not be included in the enrollment (as he/she is an unknown individual not enrolled yet). Since verification is the process of verifying an individual identity, or one-to-one matching, the target subject will have to verify himself/herself as an individual who is available in the enrollment. This should end with a rejection from the system as the target subject is an impostor and not the person he/she said to be.

The current neural network used in my research is able to be rearranged and act as a verification-oriented neural network. To do so, the target subject have to be first omitted from the training set and enrollment. When the target subject is being used as a probe, and claimed to be an individual in the enrollment, the other data within the enrollment should be changed and act as an impostor data. For example, the target subject claims to be Subject A in the enrollment, then all of the other subjects in the enrollment should change their label and combine together to be an "impostor". By doing so, the target subject will be only compared between two different pairs, that is Subject A (who the target subject claimed to be) or "impostor" (enrollment data that is not Subject A). Since there will be more variations in the classification of "impostor" trained data, the target subject is expected to be estimated as an impostor instead of Subject A. This has been illustrated in Figure 5.1.

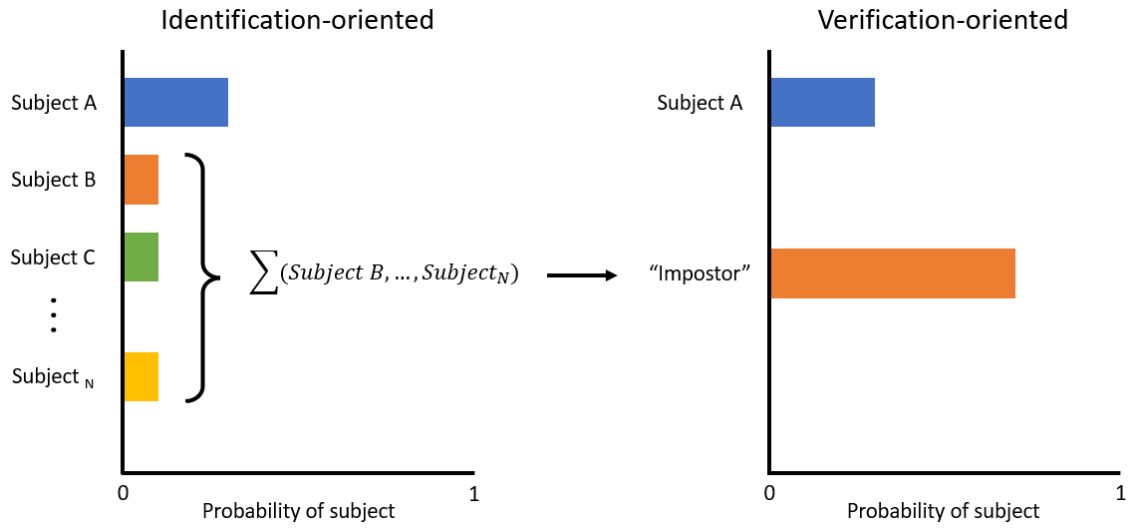


Figure 5.1: Output of current identification-oriented to verification-oriented

Another method for verification is by implementing an acceptance threshold to the probability. If the target subject's probability is over the acceptance threshold of the claimed subject, then the target subject will be accepted as genuine.

There is another neural network that can be used as verification-oriented neural network which is Siamese network [91], [92]. Siamese network contains two or more identical subnetworks consisting the same configuration with the same parameters and weights. Siamese network will use two inputs, one from the enrollment and one from the probe, and process their feature vectors. These feature vectors will then be calculated by loss functions, such as contrastive loss, to find the similarity or dissimilarity using distance metric. Another type of Siamese network uses triplet loss function [93] which requires three inputs, an anchor, a positive, and a negative. This can be viewed similarly to the verification-oriented neural network that has been previously discussed. The anchor is the probe to be tested, the positive is the claimed subject in enrollment, and the negative is any of the enrollment that is not the claimed subject in enrollment. If the anchor and the positive has a smaller distance than the distance of anchor and the negative, then the anchor will be deemed as genuine, and vice versa.

In rhythm recognition of one-time rhythm, the limitation of cross-correlation has also been stated. One approach that can solve the overlapping problem that has been stated is by using normalized cross-correlation. The formula of normalized cross-correlation is as follows:

$$I_{NCC}(x) = \frac{\sum_{\xi=1}^{m}(f_A(\xi - x) - \bar{f}_A)(f_B(\xi) - \bar{f}_B)}{\sqrt{\sum_{\xi=1}^{m}(f_A(\xi - x) - \bar{f}_A)^2}\sqrt{\sum_{\xi=1}^{m}(f_B(\xi - x) - \bar{f}_B)^2}}$$

The formula shows that the normalized cross-correlation is calculated by signal subtracting the mean of signal then divided by the standard deviation. The value of $I_{NCC}(x)$ will range from −1 to 1, with 1 indicating perfect correlation and –1 indicating perfect anti-correlation. In normalized cross-correlation, the signal will be transformed at the y-axis where the signal mean will be at 0 of y-axis, while the amplitude of the signal will be normalized to be between 1 and –1 based on the current frame's mean and standard deviation that is being calculated. Due to this normalization, the signal does not have to be converted into binary signal as has been proposed in the research.

Also, the current sample size of participants is considered insufficient for public use, but can be sufficient for small group of individuals, such as for local network access, private office access, and other applications for private use or small group. In addition, the current Myo Armband has a limit sampling rate of 200 Hz, which can have difficult to pick up high frequency or high BPM musical notes. This can be improved by using other EMG device.

References

[1] T. Matsumoto and H. Matsumoto, “Impact of artificial gummy fingers on fingerprint systems,” *Opt. Secur. Counterfeit Deterrence Tech. IV*, vol. 4677, no. May, pp. 275–289, 2002, [Online]. Available: https://doi.org/10.1117/12.462719

[2] R. Bolle, S. Pankanti, and A. K. Jain, *Biometrics, Personal Identification in Networked Society: Personal Identification in Networked Society*. USA: Kluwer Academic Publishers, 1998.

[3] R. Clarke, “Human Identification in Information Systems: Management Challenges and Public Policy Issues,” *Inf. Technol. People*, vol. 7, no. 4, pp. 6–37, 1994, doi: 10.1108/09593849410076799.

[4] Apple Inc., “About Face ID advanced technology,” 2022. https://support.apple.com/en-us/HT208108 (accessed May 18, 2022).

[5] S. Rahmatalla, H. J. Kim, M. Shanahan, and C. C. Swan, “Effect of restrictive clothing on balance and gait using motion capture and stability analysis,” *SAE Tech. Pap.*, vol. 114, no. May 2022, pp. 713–722, 2005, doi: 10.4271/2005-01-2688.

[6] C. S. and de S. S. A. and del P. G. B. and V. V. J. Casanova J. Guerraand Ávila, “A Real-Time In-Air Signature Biometric Technique Using a Mobile Device Embedding an Accelerometer,” in *Networked Digital Technologies*, 2010, pp. 497–503.

[7] R. Dhamija and A. Perrig, “Deja Vu–A User Study: Using Images for Authentication,” Aug. 2000. [Online]. Available: https://www.usenix.org/conference/9th-usenix-security-symposium/deja-vu-user-study-using-images-authentication

[8] L. O’Gorman, “Comparing passwords, tokens, and biometrics for user authentication,” in *Proceedings of the IEEE*, 2003, vol. 91, no. 12, pp. 2021–2040. doi: 10.1109/JPROC.2003.819611.

[9] K. Abhishek, S. Roshan, P. Kumar, and R. Ranjan, “A Comprehensive Study on Multifactor Authentication Schemes,” in *Advances in Intelligent Systems and Computing*, vol. 177 AISC, no. VOL. 2, 2013, pp. 561–568. doi: 10.1007/978-3-642-31552-7_57.

[10] N. Harini and T. R. Padmanabhan, “2CAuth: A new two factor authentication scheme using QR-code,” *Int. J. Eng. Technol.*, vol. 5, no. 2, pp. 1087–1094, 2013.

[11] A. Ometov, S. Bezzateev, N. Mäkitalo, S. Andreev, T. Mikkonen, and Y. Koucheryavy, “Multi-Factor Authentication: A Survey,” *Cryptography*, vol. 2, no. 1, pp. 1–31, 2018, doi: 10.3390/cryptography2010001.

[12] H. Kim, J. Han, C. Park, and O. Yi, “Analysis of vulnerabilities that can occur when generating one-time password,” *Appl. Sci.*, vol. 10, no. 8, 2020, doi: 10.3390/APP10082961.

[13] S. Ma *et al.*, “An empirical study of SMS one-time password authentication in android apps,” in *ACM International Conference Proceeding Series*, Dec. 2019, pp. 339–354. doi: 10.1145/3359789.3359828.

[14] A. Oluwakemi Christiana, A. Noah Oluwatobi, G. Ayomide Victory, and O. Roseline Oluwaseun, “A Secured One Time Password Authentication Technique using (3, 3) Visual Cryptography Scheme,” in *Journal of Physics: Conference Series*, Oct. 2019, vol. 1299, no. 1. doi: 10.1088/1742-6596/1299/1/012059.

[15] D.C. Hitchcock, “Evaluation and Combination of Biometric Authentication,” University of Florida, 2003.

[16] W. Yang, S. Wang, J. Hu, G. Zheng, and C. Valli, “Security and accuracy of fingerprint-based biometrics: A review,” *Symmetry (Basel).*, vol. 11, no. 2, 2019, doi: 10.3390/sym11020141.

[17] V. Rajasekar *et al.*, “Enhanced multimodal biometric recognition approach for smart cities based on an optimized fuzzy genetic algorithm,” *Sci. Rep.*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-021-04652-3.

[18] C. Mitchell and C.-C. Shing, “Discussing Alternative Login Methods and Their Advantages and Disadvantages,” in *2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, 2018, pp. 1353–1356. doi: 10.1109/FSKD.2018.8687163.

[19] R. Janakiraman and T. Sim, “Keystroke dynamics in a general setting,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 4642 LNCS, pp. 584–593, 2007, doi: 10.1007/978-3-540-74549-5_62.

[20] J. Ho and D. K. Kang, “Sequence alignment of dynamic intervals for keystroke dynamics based user authentication,” *2014 Jt. 7th Int. Conf. Soft Comput. Intell. Syst. SCIS 2014 15th Int. Symp. Adv. Intell. Syst. ISIS 2014*, pp. 1433–1438, 2014, doi: 10.1109/SCIS-ISIS.2014.7044658.

[21] P. Maharjan *et al.*, “Keystroke Dynamics based Hybrid Nanogenerators for Biometric Authentication and Identification using Artificial Intelligence,” *Adv. Sci.*, vol. 8, no. 15, pp. 1–9, 2021, doi: 10.1002/advs.202100711.

[22] B. Choudhury, P. Then, B. Issac, V. Raman, and M. Haldar, “A Survey on Biometrics and Cancelable Biometrics Systems,” *Int. J. Image Graph.*, vol. 18, p. 1850006, Jun. 2018, doi: 10.1142/S0219467818500067.

[23] V. M. Patel, N. K. Ratha, and R. Chellappa, “Cancelable biometrics: A review,” *IEEE Signal Processing Magazine*, vol. 32, no. 5. Institute of Electrical and Electronics Engineers Inc., pp. 54–65, Sep. 01, 2015. doi: 10.1109/MSP.2015.2434151.

[24] Manisha and N. Kumar, “Cancelable Biometrics: a comprehensive survey,” *Artif.*

Intell. Rev., vol. 53, no. 5, pp. 3403–3446, Jun. 2020, doi: 10.1007/s10462-019-09767-8.

[25] M. Damasceno and A. M. P. Canute, "An empirical analysis of ensemble systems in cancellable behavioural biometrics: A touch screen dataset," in *Proceedings of the International Joint Conference on Neural Networks*, Sep. 2014, pp. 2661–2668. doi: 10.1109/IJCNN.2014.6889819.

[26] M. Damasceno and A. M. P. Canuto, "An empirical analysis of cancellable transformations in a behavioural biometric modality," in *13th International Conference on Hybrid Intelligent Systems, HIS 2013*, Oct. 2014, pp. 149–154. doi: 10.1109/HIS.2013.6920473.

[27] P. S. Teh, "USING USERS' TOUCH DYNAMICS BIOMETRICS TO ENHANCE AUTHENTICATION ON MOBILE DEVICES," University of Manchester, Manchester, England, 2019.

[28] E. Ellavarason, R. Guest, F. Deravi, R. Sanchez-Riello, and B. Corsetti, "Touch-dynamics based Behavioural Biometrics on Mobile Devices -A Review from a Usability and Performance Perspective," *ACM Computing Surveys*, vol. 53, no. 6. Association for Computing Machinery, Feb. 01, 2021. doi: 10.1145/3394713.

[29] O. C. Hayes, "The Use of Melodic and Rhythmic MnemonicsTo Improve Memory and Recall in Elementary Students in the Content AreasACKNOWLEDGEMENTS," Dominican University of California, San Rafael, 2009.

[30] T. Hartley, M. J. Hurlstone, and G. J. Hitch, "Effects of rhythm on memory for spoken sequences: A model and tests of its stimulus-driven mechanism," *Cogn. Psychol.*, vol. 87, pp. 135–178, Jun. 2016, doi: 10.1016/j.cogpsych.2016.05.001.

[31] Daniel Müllensiefen, "Musical Memory," Apr. 02, 2019. https://serious-science.org/musical-memory-9412 (accessed Jun. 12, 2022).

[32] S. A. El-Moneim *et al.*, "Text-dependent and text-independent speaker recognition of reverberant speech based on CNN," *Int. J. Speech Technol.*, vol. 24, no. 4, pp. 993–1006, Dec. 2021, doi: 10.1007/s10772-021-09805-3.

[33] E. S. Shimaa Zeid, R. A. ElKamar, and S. I. Hassan, "Fixed-Text vs. Free-Text Keystroke Dynamics for User Authentication," 2022.

[34] Catherine Matacic, "Rhythm might be hardwired in humans," *Science*, Dec. 19, 2016. https://www.science.org/content/article/rhythm-might-be-hardwired-humans (accessed Jun. 08, 2022).

[35] F. L. Bouwer, C. M. Werner, M. Knetemann, and H. Honing, "Disentangling beat perception from sequential learning and examining the influence of attention and musical abilities on ERP responses to rhythm," *Neuropsychologia*, vol. 85, pp. 80–90, 2016, doi: 10.1016/j.neuropsychologia.2016.02.018.

[36] F. L. Bouwer, "What do we need to hear a beat? The influence of attention, musical abilities, and accents on the perception of metrical rhythm," University of Amsterdam, 2016.

[37] R. H. Chowdhury, M. B. I. Reaz, M. A. Bin Mohd Ali, A. A. A. Bakar, K. Chellappan, and T. G. Chang, “Surface electromyography signal processing and classification techniques,” *Sensors (Switzerland)*, vol. 13, no. 9. MDPI AG, pp. 12431–12466, Sep. 17, 2013. doi: 10.3390/s130912431.

[38] P. Visconti, F. Gaetani, G. A. Zappatore, and P. Primiceri, “Technical features and functionalities of Myo armband: An overview on related literature and advanced applications of myoelectric armbands mainly focused on arm prostheses,” *Int. J. Smart Sens. Intell. Syst.*, vol. 11, no. 1, pp. 1–25, 2018, doi: 10.21307/ijssis-2018-005.

[39] J. R. Schofield, “Electrocardiogram Signal Quality Comparison Between a Dry Electrode and a Standard Wet Electrode over a Period of Extended Wear,” CLEVELAND STATE UNIVERSITY, 2012. [Online]. Available: https://engagedscholarship.csuohio.edu/cgi/viewcontent.cgi?article=1714&context=etdarchive

[40] P. Laferriere, E. D. Lemaire, and A. D. C. Chan, “Surface electromyographic signals using dry electrodes,” *IEEE Trans. Instrum. Meas.*, vol. 60, no. 10, pp. 3259–3268, 2011, doi: 10.1109/TIM.2011.2164279.

[41] J. C. Ives and J. K. Wigglesworth, “Sampling rate effects on surface EMG timing and amplitude measures,” *Clin. Biomech.*, vol. 18, no. 6, pp. 543–552, 2003, doi: 10.1016/S0268-0033(03)00089-5.

[42] A. Phinyomark and E. Scheme, “A feature extraction issue for myoelectric control based on wearable EMG sensors,” *2018 IEEE Sensors Appl. Symp. SAS 2018 - Proc.*, vol. 2018-Janua, pp. 1–6, 2018, doi: 10.1109/SAS.2018.8336753.

[43] Reality Labs, “Inside Facebook Reality Labs: The next era of human-computer interaction,” Mar. 09, 2021. https://tech.fb.com/ar-vr/2021/03/inside-facebook-reality-labs-the-next-era-of-human-computer-interaction/ (accessed Jun. 08, 2022).

[44] Reality Labs, “Inside Facebook Reality Labs: Wrist-based interaction for the next computing platform,” Mar. 18, 2021. https://tech.fb.com/ar-vr/2021/03/inside-facebook-reality-labs-wrist-based-interaction-for-the-next-computing-platform/ (accessed Jun. 08, 2022).

[45] A. M. H. Wong and D. K. Kang, “Stationary Hand Gesture Authentication Using Edit Distance on Finger Pointing Direction Interval,” *Sci. Program.*, vol. 2016, 2016, doi: 10.1155/2016/7427980.

[46] S. Kiranyaz, T. Ince, O. Abdeljaber, O. Avci, and M. Gabbouj, “1-D Convolutional Neural Networks for Signal Processing Applications,” in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 8360–8364. doi: 10.1109/ICASSP.2019.8682194.

[47] S. A. Raurale, J. McAllister, and J. M. Del Rincon, “EMG Biometric Systems Based on Different Wrist-Hand Movements,” *IEEE Access*, vol. 9, pp. 12256–12266, 2021, doi: 10.1109/ACCESS.2021.3050704.

[48] J. He and N. Jiang, “Biometric From Surface Electromyogram (sEMG): Feasibility of User Verification and Identification Based on Gesture Recognition,” *Front. Bioeng. Biotechnol.*, vol. 8, Feb. 2020, doi: 10.3389/fbioe.2020.00058.

[49] J. Lien *et al.*, “Soli: Ubiquitous gesture sensing with millimeter wave radar,” *ACM Trans. Graph.*, vol. 35, no. 4, 2016, doi: 10.1145/2897824.2925953.

[50] S. Ahmed, K. D. Kallu, S. Ahmed, and S. H. Cho, “Hand gestures recognition using radar sensors for human-computer-interaction: A review,” *Remote Sensing*, vol. 13, no. 3. MDPI AG, pp. 1–24, Feb. 01, 2021. doi: 10.3390/rs13030527.

[51] R. J. Elble, “Gravitational artifact in accelerometric measurements of tremor,” *Clin. Neurophysiol.*, vol. 116, no. 7, pp. 1638–1643, Jul. 2005, doi: 10.1016/j.clinph.2005.03.014.

[52] F. Xu *et al.*, “Recent developments for flexible pressure sensors: A review,” *Micromachines*, vol. 9, no. 11. MDPI AG, Nov. 07, 2018. doi: 10.3390/mi9110580.

[53] A. S. Sadun, J. Jalani, and J. A. Sukor, “Force Sensing Resistor (FSR): a brief overview and the low-cost sensor for active compliance control,” in *First International Workshop on Pattern Recognition*, Jul. 2016, vol. 10011, p. 1001112. doi: 10.1117/12.2242950.

[54] G. Li, D. Li, Y. Cheng, W. Sun, X. Han, and C. Wang, “Design of pressure-sensing diaphragm for MEMS capacitance diaphragm gauge considering size effect,” *AIP Adv.*, vol. 8, no. 3, Mar. 2018, doi: 10.1063/1.5021374.

[55] S. M. Khan, R. B. Mishra, N. Qaiser, A. M. Hussain, and M. M. Hussain, “Diaphragm shape effect on the performance of foil-based capacitive pressure sensors,” *AIP Adv.*, vol. 10, no. 1, Jan. 2020, doi: 10.1063/1.5128475.

[56] M. B. I. Reaz, M. S. Hussain, and F. Mohd-Yasin, “Techniques of EMG signal analysis: Detection, processing, classification and applications,” *Biol. Proced. Online*, vol. 8, no. 1, pp. 11–35, Mar. 2006, doi: 10.1251/bpo115.

[57] M. Niwa *et al.*, “Detection and Transmission of" Tsumori": an Archetype of Behavioral Intention in Controlling a Humanoid Robot,” *20th Int. Conf. Artif. Real. Telexistence*, pp. 193–196, 2010, [Online]. Available: http://icat.vrsj.org/ICAT2010_Proceedings/poster/paper-107.pdf%0Ahttp://www.vrsj.org/ic-at/ICAT2010_Proceedings/poster/paper-107.pdf

[58] K. Ahmadian and M. Gavrilova, “Chaotic Neural Network for Biometric Pattern Recognition,” *Adv. Artif. Intell.*, vol. 2012, pp. 1–9, Aug. 2012, doi: 10.1155/2012/124176.

[59] M. Kraus, S. Feuerriegel, and A. Oztekin, “Deep learning in business analytics and operations research: Models, applications and managerial implications,” *Eur. J. Oper. Res.*, vol. 281, no. 3, pp. 628–641, Mar. 2020, doi: 10.1016/j.ejor.2019.09.018.

[60] S. Sakunthala, R. Kiranmayi, and P. N. Mandadi, “A review on artificial intelligence techniques in electrical drives: Neural networks, fuzzy logic, and genetic algorithm,” in *2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon)*, 2017, pp. 11–16. doi: 10.1109/SmartTechCon.2017.8358335.

[61] Aravindpai Pai, “CNN vs. RNN vs. ANN – Analyzing 3 Types of Neural Networks in Deep Learning,” Feb. 17, 2020. https://www.analyticsvidhya.com/blog/2020/02/cnn-vs-rnn-vs-mlp-analyzing-3-types-of-neural-networks-in-deep-learning/ (accessed Jun. 26, 2022).

[62] A. Jalaly Bidgoly, H. Jalaly Bidgoly, and Z. Arezoumand, “A survey on methods and challenges in EEG based authentication,” *Computers and Security*, vol. 93. Elsevier Ltd, Jun. 01, 2020. doi: 10.1016/j.cose.2020.101788.

[63] N. Sharma, S. Gupta, and P. Mehta, “A Comprehensive Study on Offline Signature Verification,” in *Journal of Physics: Conference Series*, Jul. 2021, vol. 1969, no. 1. doi: 10.1088/1742-6596/1969/1/012044.

[64] S. Sigtia, E. Marchi, S. Kajarekar, D. Naik, and J. Bridle, “Multi-task Learning for Speaker Verification and Voice Trigger Detection,” Jan. 2020, doi: 10.1109/ICASSP40776.2020.9054760.

[65] Z. Fan, M. Li, S. Zhou, and B. Xu, “Exploring wav2vec 2.0 on speaker verification and language identification,” 2021.

[66] Z. Ma, H. Yu, W. Chen, and J. Guo, “Short utterance based speech language identification in intelligent vehicles with time-scale modifications and deep bottleneck features,” *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 121–128, Jan. 2019, doi: 10.1109/TVT.2018.2879361.

[67] O. Krčadinac, U. Šošević, and D. Starčević, “Evaluating the performance of speaker recognition solutions in e-commerce applications,” *Sensors*, vol. 21, no. 18, Sep. 2021, doi: 10.3390/s21186231.

[68] C. S. Vorugunti, R. K. S. Gorthi, and V. Pulabaigari, “Online signature verification by few-shot separable convolution based deep learning,” in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, Sep. 2019, pp. 1125–1130. doi: 10.1109/ICDAR.2019.00182.

[69] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia, “DeepSign: Deep On-Line Signature Verification,” *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. PP. Jun. 2021. doi: 10.1109/TBIOM.2021.3054533.

[70] C. S. V., Anoushka, P. Mukherjee, and V. P., “A Light Weight and Hybrid Deep Learning Model Based Online Signature Verification,” in *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, 2019, vol. 5, pp. 53–58. doi: 10.1109/ICDARW.2019.40081.

[71] J. Fierrez Julianand Ortega-Garcia, “On-Line Signature Verification,” in *Handbook of Biometrics*, P. and R. A. A. Jain Anil K.and Flynn, Ed. Boston,

MA: Springer US, 2008, pp. 189–209. doi: 10.1007/978-0-387-71041-9_10.

[72] C. S. Vorugunti, P. Mukherjee, D. S. Guru, and V. Pulabaigari, “Online Signature Verification Based on Writer Specific Feature Selection and Fuzzy Similarity Measure,” 2019.

[73] M. B. Yilmaz and K. Öztürk, “Hybrid User-Independent and User-Dependent Offline Signature Verification with a Two-Channel CNN,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018, pp. 639–6398. doi: 10.1109/CVPRW.2018.00094.

[74] H. Zeinali, B. BabaAli, and H. Hadian, “Online signature verification using I-vector representation,” *IET Biometrics*, vol. 7, no. 5, pp. 405–414, Sep. 2018, doi: 10.1049/iet-bmt.2017.0059.

[75] A. Mccabe and G. Gupta, “A Review of Dynamic Handwritten Signature Verification A Review of Dynamic Handwritten Signature Veriication,” 1998. [Online]. Available: https://www.researchgate.net/publication/2352171

[76] S. Sprager and M. B. Juric, “Inertial sensor-based gait recognition: A review,” *Sensors (Switzerland)*, vol. 15, no. 9. MDPI AG, pp. 22089–22127, Sep. 02, 2015. doi: 10.3390/s150922089.

[77] L. Rong, Z. Jianzhong, L. Ming, and H. Xiangfeng, “A Wearable Acceleration Sensor System for Gait Recognition,” in *2007 2nd IEEE Conference on Industrial Electronics and Applications*, 2007, pp. 2654–2659. doi: 10.1109/ICIEA.2007.4318894.

[78] D. Gafurov and E. Snekkenes, “Gait recognition using wearable motion recording sensors,” *EURASIP J. Adv. Signal Process.*, vol. 2009, 2009, doi: 10.1155/2009/415817.

[79] M. O. Derawi, P. Bours, and K. Holien, “Improved cycle detection for accelerometer based gait authentication,” in *Proceedings - 2010 6th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IIHMSP 2010*, 2010, pp. 312–317. doi: 10.1109/IIHMSP.2010.84.

[80] Roy Soumen, Roy Utpal, and Sinha D. D., “Security Enhancement of Knowledge-based User Authentication through Keystroke Dynamics,” *MATEC Web Conf.*, vol. 57, p. 2004, 2016, doi: 10.1051/matecconf/20165702004.

[81] H. Çeker and S. Upadhyaya, “Sensitivity analysis in keystroke dynamics using convolutional neural networks,” in *2017 IEEE Workshop on Information Forensics and Security (WIFS)*, 2017, pp. 1–6. doi: 10.1109/WIFS.2017.8267667.

[82] P. S. Teh, A. B. J. Teoh, and S. Yue, “A survey of keystroke dynamics biometrics,” *The Scientific World Journal*, vol. 2013. Hindawi Publishing Corporation, 2013. doi: 10.1155/2013/408280.

[83] A. Acien, J. V. Monaco, A. Morales, R. Vera-Rodriguez, and J. Fierrez, “TypeNet: Scaling up Keystroke Biometrics,” Apr. 2020, [Online]. Available: http://arxiv.org/abs/2004.03627

[84] M. Pleva, P. Bours, S. Ondáš, and J. Juhár, “Improving static audio keystroke analysis by score fusion of acoustic and timing data,” *Multimed. Tools Appl.*, vol. 76, no. 24, pp. 25749–25766, Dec. 2017, doi: 10.1007/s11042-017-4571-7.

[85] J. Roth, X. Liu, A. Ross, and D. Metaxas, “Biometric authentication via keystroke sound,” in *2013 International Conference on Biometrics (ICB)*, 2013, pp. 1–8. doi: 10.1109/ICB.2013.6613015.

[86] F. Okumura, A. Kubota, Y. Hatori, K. Matsuo, M. Hashimoto, and A. Koike, “A Study on Biometric Authentication based on Arm Sweep Action with Acceleration Sensor,” in *2006 International Symposium on Intelligent Signal Processing and Communications*, 2006, pp. 219–222. doi: 10.1109/ISPACS.2006.364871.

[87] J. Liu, Z. Wang, L. Zhong, J. Wickramasuriya, and V. Vasudevan, “uWave: Accelerometer-based Personalized Gesture Recognition and Its Applications,” 2009.

[88] C. Shen, Z. Wang, C. Si, Y. Chen, and X. Su, “Waving gesture analysis for user authentication in the mobile environment,” *IEEE Netw.*, vol. 34, no. 2, pp. 57–63, Mar. 2020, doi: 10.1109/MNET.001.1900184.

[89] J. Guerra-Casanova, C. Sánchez-Ávila, G. Bailador, and A. de Santos Sierra, “Authentication in mobile devices through hand gesture recognition,” *Int. J. Inf. Secur.*, vol. 11, no. 2, pp. 65–83, Apr. 2012, doi: 10.1007/s10207-012-0154-9.

[90] T. Derrick, “Time Series Analysis: The Cross-Correlation Function,” in *Innovative analyses of Human Movement*, 2004, pp. 189–205.

[91] G. Koch, R. Zemel, R. Salakhutdinov, and others, “Siamese neural networks for one-shot image recognition,” in *ICML deep learning workshop*, 2015, vol. 2, p. 0.

[92] A. Abbas and D. Moser, “Siamese Network Training Using Sampled Triplets and Image Transformation,” *ArXiv*, vol. abs/2106.07015, 2021.

[93] F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering,” Mar. 2015, doi: 10.1109/CVPR.2015.7298682.

Acknowledgement

I would like to express my sincere gratitude to all the people who have been accompanying me during my study in Osaka University. I would not have been able to complete this dissertation by myself without their help. I would like to thank each and every one of them.

First and foremost, I would like to express my special appreciation and thanks to my advisors, Professor Taro Maeda and Associate Professor Masahiro Furukawa. They have been tremendous mentors for me. I would like to thank them for encouraging me in my research and for guiding me in becoming a research scientist. Their advice on both research as well as on my career have been priceless.

I would like to thank Professor Hideo Matsuda and Professor Yasushi Makihara for their comments and remarks in refining and improving the quality of my dissertation.

I am also grateful to all my friends in Osaka University who have been giving me a helpful hand during my years; especially my lab mates, Hiroki Miyamoto, Makoto Aoto, Naoki Nishio, Junjie Hua, Akihiro Terashima, Masataka Kurokawa, and Teppei Matsuoka who have given me support and advice on my research and Japanese culture.

Lastly, I would like to give thanks to my family, especially my spouse, Laura Tiong Siew Zin. Their belief has kept my spirits and motivation high during this process. An extended gratitude to my spouse who move to Japan to be by my side.