**Development of a Mobile Device-based Augmented Reality Mirror Therapy System for Rehabilitation of the Stroke Hemiplegia**

Po-Yu Huang[[1]](#footnote-0)\* Che-Wei Lin [[2]](#footnote-1)\*\*

Department of Biomedical Engineering

National Cheng Kung University, Tainan 701, Taiwan, R.O.C.

Abstract

This study developed an augmented reality mirror therapy (ARMT) system based on mobile phones, aiming to provide the convenience of traditional mirror therapy (MT) while providing a relatively good rehabilitation quality. Ten young healthy subjects were recruited to participate in the clinical trial and divided into two groups. In the group, traditional MT and ARMT interventions were performed to observe the difference in hand function before and after the intervention. Functional near infrared spectroscopy (fNIRS) was used to evaluate the activation of different brain regions during the intervention. The results of the study found that there were significant differences in the pre- and post-tests of ARMT, but not in MT; and found that the two had similar blood perfusion patterns, which indicated the potential of ARMT in clinical rehabilitation applications.

Catalog

[Abstract 1](#_Toc15048)

[Chapter 1 Overview 4](#_Toc9489)

[1.1 INTRODUCTION 4](#_Toc16969)

[1.1.1 Background 4](#_Toc13788)

[1.1.2 Mirror Therapy (MT) 5](#_Toc6940)

[1.1.3 Virtual Reality Therapy 5](#_Toc1631)

[1.1.4 Functional electrical stimulation (FES) 6](#_Toc8897)

[1.1.5 Trend of Digital Health 6](#_Toc25996)

[1.1.6 Digital Therapeutics 7](#_Toc32261)

[1.1.7 Telemedicine 8](#_Toc22527)

[1.2 LITERATURE REVIEW 9](#_Toc19983)

[1.2.1 Rehabilitation on Stroke Patient 9](#_Toc21844)

[1.2.2 Telerehabilitation 10](#_Toc25304)

[1.2.3 Digital Rehabilitation 11](#_Toc28995)

[1.2.4 The combination of Virtual/Augmented Reality 11](#_Toc32336)

[1.2.5 Immersive Rehabilitation via Mobile Phone 12](#_Toc22189)

[1.2.6 Quantitative Analysis on Mirror Therapy 13](#_Toc5567)

[1.3 OBJECTIVE 14](#_Toc27641)

[1.3.1 Objective of This Study 14](#_Toc11049)

[Chapter 2 Methodology and Material 15](#_Toc29555)

[2.1 SYSTEM DEVELOPMENT 15](#_Toc26290)

[2.1.1 Augmented Reality Mirror Therapy System (ARMT) 15](#_Toc20367)

[2.1.2 Hardware and Accessories 16](#_Toc6238)

[2.1.3 System Architecture 16](#_Toc4657)

[2.2 DEVELOPMENT PROGRESS 17](#_Toc19619)

[2.2.1 Development Tools 17](#_Toc30586)

[2.2.2 Hand Joints Skeleton Approach 18](#_Toc12960)

[2.2.3 Hand Contour Approach 19](#_Toc12191)

[2.2.4 Cardboard and Screen Fluency Approach 19](#_Toc18681)

[2.3 EXPERIMENTAL DESIGN 20](#_Toc14551)

[2.3.1 Purpose 20](#_Toc11943)

[2.3.2 Participant Criteria 20](#_Toc20841)

[2.3.3 Procedure 20](#_Toc8000)

[[2.3.4] Hand Function Assessment Tool 21](#_Toc2899)

[[2.3.5] Brain Area Activity Measurement 21](#_Toc7809)

[Chapter 4 Experimental Result 22](#_Toc5682)

[4.1 Procedure 22](#_Toc13599)

[[4.2] Participant Criteria 22](#_Toc14665)

[4.3 Enrollment Motion 22](#_Toc28970)

[4.4 Result and Comparison 22](#_Toc1753)

[[4.4.1] Hand Function Evaluation 22](#_Toc11914)

[[4.4.2] fNIRS of ROI 23](#_Toc23339)

[Chapter 5 Discussion and Conclusion 24](#_Toc4188)

[5.1 Discussion 24](#_Toc3323)

[5.2 ARMT Potential on Stroke Rehabilitation 24](#_Toc6712)

[[5.3] Conclusion 24](#_Toc8739)

[5.4 Limitation 25](#_Toc4264)

[5.5 Future Work 25](#_Toc12985)

# Chapter 1 Overview

## INTRODUCTION

## 1.1.1 Background

For post-stroke survivors, the most common sequelae are hemiplegia which is a type of paralysis that affects one side of the body, cause a range of physical impairments, including weakness, spasticity, and loss of sensation, and significantly impact a person's ability to perform daily activities. Stroke can cause a decrease in functionality in various ways, depending on the affected brain region. However, the most common and frustrating defects are related to motor and sensory impairments, which often occur in groups of survivors. This is because the motor and sensory fibers are in the superficial layer of the brain (cortex), which is more susceptible to damage from a stroke. Hence, upper limb rehabilitation is a critical practice for stroke survivors to regain functional independence and improve their quality of life. Including grasp, reach, dexterity, and coordination, the importance of upper limb rehabilitation is such that many routine rehabilitation exercises mostly include their program.

Reports from clinical and therapist experiences, most scholars agree that rehabilitation after stroke has a decisive impact on patients' prognosis. Rehabilitation quality, immediacy, duration, and patients' adherence are crucial factors that directly affect post-stroke survivors' maximum recovery potential. Rehabilitation aims to help stroke patients regain their general life skills. Upper limb rehabilitation, which focuses on recovering patients' arm, shoulder, hand, and wrist, is an essential part of the overall rehabilitation process. Not only is it complex to control, but human upper limbs also play a critical role in daily activities, such as self-care, intake, dressing, and working. The early and intensive upper limb rehabilitation can improve outcomes for stroke survivors, including better functional outcomes and restore the life quality before the stroke to the greatest extent. Hence, upper limb rehabilitation is considered a critical component of stroke rehabilitation, and it is often prioritized over other forms of rehabilitation.

Theoretically, based on the systematic rehabilitation schedule and consistent practice with a therapist, most minor stroke patients can achieve complete recovery within six months. Although extensive stroke patients may require several years and may not fully recover, early exposure to therapy is critical during the recovery process. Therefore, balancing the quality of therapy and creating an interesting and immersive experience to support their willingness to participate has become a significant topic of research.

In the recent years, there has been growing interest in innovative rehabilitation methods that incorporate technology, such as virtual reality and robotics, to enhance the effectiveness and efficiency of stroke rehabilitation. These methods offer potential advantages such as increased patient engagement, personalized feedback, and improved outcomes.

Canadian Stroke Best Practice Recommendations lists several rehabilitation methods that have a lot of evidence to support their effectiveness [[ref]](https://pubmed.ncbi.nlm.nih.gov/31983296/): functional electrical stimulation, mirror therapy and virtual reality therapy. Whether these upper limb rehabilitation methods are implemented in the early or late stage of stroke, they have obvious curative effects in more than two randomized controlled clinical trials.

## 1.1.2 Mirror Therapy (MT)

Mirror therapy is a type of therapy that is used to release phantom limb pain and improve motor function in stroke patients, especially suitable for stroke patients with hemiplegia. In a traditional mirror therapy treatment, the therapist will ask the patient sitting in front of the mirror box and hiding their affected hand into that mirror box, on the other healthy hand will be asked to place outside of the mirror box, by adjusting the angle between the patient's perspective and the mirror on the mirror box, a affected hand image reflected by the healthy hand will show up in the mirror and create a visual illusion that tricks the brain into thinking that the missing or injured limb is functioning normally, performing the same movement with the healthy hand.

After a stroke, patients may experience a phenomenon called learned non-use, where they avoid using their affected limb because it feels weak or unresponsive. This can lead to further deterioration of motor function and make it more difficult to recover. By using a mirror to create an illusion of movement in the affected limb, mirror therapy can help to overcome learned non-use and promote recovery of motor function. The visual feedback provided by the mirror can help to retrain the brain and encourage patients to use their affected limb more actively. For a treatment based on visual deception, its efficacy is believed to be related to mirror neurons and neuroplasticity principles derived from the human brain.

The basis of mirror therapy is based on the neuralplasticity of the brain which can be repaired by itself. If the patient after the stroke keeps looking at the weak side of the hand, and constantly inputting negative information to the brain, it will not be conducive to the repair of the brain. Therefore, the mirror is used to reflect the unaffected hand, so that the patient sees the image of the unaffected hand overlapping the affected side of the hand and establishes the visual illusion of the affected side in the brain.

## 1.1.3 Virtual Reality Therapy

Virtual reality (VR) therapy is a type of digital therapeutic that will be interpreted in subsection **1.2.4**. The therapy uses virtual reality technology to simulate real world-like environments to help patients overcome a variety of physical, emotional, or psychological challenges. VR therapy can be used in a range of settings, including occupational therapy, physical therapy, and mental health therapy. Depending on the use case, research in VR therapy is broadly divided into immersive VR therapy and non-immersive VR therapy. Immersive VR therapy typically involves a headset/goggle that immerses the user in a computer-generated environment. non-immersive VR therapy, on the other hand, uses a flat display or projection to achieve a similar effect, and the user will still have part of his visual perception exposed to the real world.

By interacting with this virtual environment in a variety of ways, and using hand-held controllers or other devices, the VR experience can be adjusted to meet the specific needs and goals of the individual, and the therapist can guide and monitor the therapy session in real-time. Most of the VR training course can be designed and developed by software engineers, and can be adjust according to different needs, and different users. A well-designed immersive VR therapy system can combine with another therapy practice, but more entertaining and fulfilling, which is an advantage for stroke patients who need to undergo rehabilitation for a long period.

## 1.1.4 Functional electrical stimulation (FES)

FES uses small electrical currents to stimulate the muscles of the affected limb, this rehabilitation method usually targeted the wrist and forearm muscles to reduce motor impairment and improve function. promoting movement and strength. Due to the signal blockade caused by the damage to the associated motor cortex, stroke patients have difficulty triggering their bone muscles because of their impaired motor control, FES can surrogate the patient's motor intention to trigger related acceptor of muscle. In current research progress, also based on the induction principle of neuroplasticity, the current mainstream hypothesis is that if the use of FES can synchronize the patient's motor intentions, there will be a better curative effect. Therefore, the FES rehabilitation systems combining electromyography (EMG) or even brain-computer interface (BCI) to predict motor intention have been proposed in recent years [[ref]](https://pubmed.ncbi.nlm.nih.gov/30763238/)[[ref]](https://pubmed.ncbi.nlm.nih.gov/30902630/).

## 1.1.5 Trend of Digital Health

According to the World Health Organization’s (WHO) definition in Global Strategy on Digital Health 2020-2025, digital health expands the concept of eHealth to include digital consumers, with a wider range of smart and connected devices. It also encompasses other uses of digital technologies for health such as the Internet of Things (IOT), advanced computing, big data analytics, artificial intelligence including machine learning, and robotics [[ref]](https://apps.who.int/iris/bitstream/handle/10665/344249/9789240020924-eng.pdf).

The concept of digital health is becoming increasingly important because it has the potential to revolutionize the way we deliver and receive healthcare services. Contrast to adopt healthcare workers, digital health refers to the use of technology to support and enhance healthcare services, including the delivery of therapy, evaluate, and scheduler system.

One of the main benefits of digital health is its ability to improve availability. With relative technologies, patients in need would have a chance to access healthcare services from anywhere, at any time, without the need for in-person visits. This is especially beneficial for stroke patients who may have difficulty accessing traditional healthcare services due to their defect on mobility or other barriers.

Digital health can also improve the healthcare quality that patients receive. With the use of digital tools such as telemedicine, remote monitoring, and online therapy platforms, healthcare providers can more easily monitor and track patients' progress, identify potential issues or complications, and adjust treatment plans as needed.

Furthermore, digital health can help to reduce healthcare costs by streamlining administrative tasks, improving efficiency, and reducing the need for in-person visits. This can lead to cost savings for both healthcare providers and patients.

In the recent years, due to its potential in clinical application, digital health has increased the rehab efficacy in the pandemic. Two of the following section, that is, digital therapeutics and telemedicine, will be proposed.

## 1.1.6 Digital Therapeutics

Digital therapeutics is an subset of digital health, they are evidence-based therapeutic interventions driven by high quality software programs to prevent, manage, or treat a medical disorder or disease [[ref]](https://www.nature.com/articles/nbt.3495). Same as the traditional therapeutics, a new proposed digital therapeutic should publish the reproducible trail result and meaningful clinical outcome to claim its effectiveness and certified by national regulations. As a innovative field of therapeutic, digital therapeutic treatment is expected to start from the patient's behavior and lifestyle habits, because of the digital nature of the methodology, data can be collected and analyzed as both a progress report and a preventative measure [[ref]](https://www.sciencedirect.com/science/article/pii/S1359644616000301?via%3Dihub).

The use of digital therapeutics to improve health outcomes dates as far back as 2000 [[ref]](https://pubmed.ncbi.nlm.nih.gov/19816750/)., and the first mention of the term in a peer-reviewed research publication was in 2015, in which Dr. Cameron Sepah formally defined the field as: "Digital therapeutics are evidence-based behavioral treatments delivered online that can increase accessibility and effectiveness of health care [[ref]](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4409647/). For now days, digital therapeutics don’t just refer to one of these medical products or methods, products driven by digital devices or software that deliver medical interventions and therapies that can be called one of the digital therapeutics.

Digital therapeutics have made surprising progress since the evolution of the edge device cause in most of the cases. One example of the application is real time physiological signal monitor, the therapy of chronic disease patients have a high dependency with their life-style and behavior changes, by utilizing low-cost devices nearby the side, physiological signal in time can be detected, collected and evaluated by patients and the medical assistants.

## 1.1.7 Telemedicine

Telemedicine is another one of a subset in digital health, it is a form of healthcare that allows healthcare providers to communicate with their patient remotely by using several communication technologies such as live streaming, phone calls, or non-instant methods like email, messaging platforms, and fax. Telemedicine is easy to perform, however, in many national laws and regulations, medical providers are only allowed to use remote consultation to issue medical orders under special circumstances. This is because of the concern that the lack of judgment information caused by distance restrictions will cause doctors to provide wrong or hasty prescriptions.

Telemedicine has been around for several years but has become increasingly popular and important during the pandemic era. The COVID-19 pandemic has caused a huge impact to traditional healthcare systems, policy lockdown and the treatment of the virus make patients difficult to transport between home and hospital. Telemedicine has provided a solution, reducing their risk of exposure to the virus while maintaining continuity of care. For patients with stable chronic diseases, the telemedicine solution can be used for routine consultation. For simple, less human-assisted routine rehabilitation needs, the telemedicine solution has the potential to be implemented. Telemedicine has provided an alternative, convenient way for patients with chronic conditions or other health issues to visit a healthcare facility with high risk and cost.

Source from Taiwan's Ministry of Health, in practice, telemedicine process design should preserve the quality in the traditional medication process and overcome some derivative problems:

1. Access: Accessibility should take into account both information and geographic construction. Without good construction, there will be no excellent medical quality.
2. Privacy: Information security issues such as where the information is stored, whether to protect the patient's personal information, etc.
3. Diagnostic Accuracy: Telemedicine is still a medical service, and diagnostic accuracy should be considered.
4. Communication: Telemedicine should also provide good communication services for people with special disabilities and language barriers.
5. Psychological & Emotional Safety: Telemedicine cannot be used for face-to-face diagnosis, and patients' sense of safety when seeking medical treatment should be more concerned.
6. System Design (Human Factors & System Design): Whether the overall process of telemedicine is integrated into the original medical service process.

Post pandemic era has revealed the potential for telemedicine to improve access to healthcare for patients in remote or underserved areas, or for those who have difficulty accessing traditional healthcare services for various reasons.

## 1.2 LITERATURE REVIEW

To confirm the principles of the existing rehabilitation methods, in the literature research, this study focuses on stroke rehabilitation based on the principle of neuroplasticity. Since stroke rehabilitation requires high-frequency implementation, the literature on home rehabilitation and digital medicine will also be included. Finally, we found that many rehabilitation assessments are limited to subjective data such as scales. We also discuss the literature that quantifies brain signal changes as an assessment method.

## 1.2.1 Rehabilitation on Stroke Patient

Many current rehabilitation treatments for hemiplegic stroke patients are grounded in the theory of neuroplasticity, which refers to the brain's ability to reorganize and grow by altering its neural connections over time [[ref]](https://books.google.com.tw/books/about/Neuroplasticity.html?id=vvjdDAAAQBAJ&redir_esc=y). This phenomenon involves rewiring the brain to function differently from its previous state.

Related research suggests that multisensory stimulation and explicit feedback principles should be implemented for motor-oriented rehabilitation. The consistency of action and motor intention, one or multiple sensory stimulation at the same time, is an important factor in inducing neuroplasticity [[ref]](https://pubmed.ncbi.nlm.nih.gov/31920570/), and this factor is also reflected in the research direction of these three effective stroke rehabilitation methods. As an indispensable role in dealing with daily life, upper limb rehabilitation training is currently the most mainstream and most researched rehabilitation goal, besides, two-thirds of the studied interventions deal with motor recovery in recent years [[ref]](https://pubmed.ncbi.nlm.nih.gov/31983296/).

Neuroplasticity is not a rare mechanism to human brain [[ref]](https://pubmed.ncbi.nlm.nih.gov/24883212/)[[ref]](https://pubmed.ncbi.nlm.nih.gov/27507957/). To keep adapting different conditions, our brain is always changing. For the post-stroke patients, the mechanism of neuroplasticity let their brain try to repair itself, but since the dramatically changes of the patients’ behavior after the trauma such as reduced use of affected side and self-limitation due to disability, the nature process of these mechanism are often limited and misleading, the main objective of the stroke rehabilitation is to correct the way that neuroplasticity has taken [[ref]](https://www.proquest.com/docview/215286106?pq-origsite=gscholar&fromopenview=true). For rehabilitation, there have several important principles:

1. Learned non-use (Use it or lose it):

Research of the Braille readers reveals the representation in the brain have high relationship with the frequency of use of a body part [[ref]](https://pubmed.ncbi.nlm.nih.gov/8517678/)[[ref]](https://pubmed.ncbi.nlm.nih.gov/12625638/), Under the influence of neuroplasticity, the corresponding cerebral cortex area responsible for sensory finger skin even expands and inhibits the sensory ability of other infrequently used finger skin. Hence, some of the rehabilitation method such as constraint-induced therapy (CIMT) has been proposed [[ref]](https://www.ahajournals.org/doi/full/10.1161/01.STR.0000143320.64953.c4).

1. The motor control is taken over by bilateral hemisphere:

Despite most of the motor control are dominant by the contralateral hemisphere, there still have weak neural pathway on corresponding ipsilateral hemisphere in humans. Although controversial, these pathways may possibly be relevant in stroke recovery [[ref]](https://pubmed.ncbi.nlm.nih.gov/11690613/). Hence, the best recovery seems to those treatment focus on reorganizing damaged hemisphere, but bilateral symmetrical arm movement training is also seeming to be helpful to stroke patient, especially for the hemiplegia. MT is one of the represent treatments, which was initially developed to assist patients with upper limb amputation in managing phantom limb sensation and pain, as evidenced by the literature. In 1995, scholars extended this method to stroke rehabilitation and observed significant improvements in joint activity, speed, and accuracy for patients with mild, moderate, and severe limb disability. These positive outcomes not only enhance upper limb function but also motivate patients to utilize their affected side more frequently in their daily activities [[ref]](https://pubmed.ncbi.nlm.nih.gov/7566144/)

1. Multisensory input can increase the effectiveness:

Human is socialized, but the study indicated that stroke patients are frequently alone for approximately 60% of their time, and during their waking hours, they engage in physical inactivity nearly 75% of the time [[ref]](https://pubmed.ncbi.nlm.nih.gov/14988574/). The ultimate goal of rehabilitation is to get patients back into daily life, including complex and stimulating social behaviors, Some studies advocate that exposing patients to multisensory stimuli, such as rehabilitation fields that include motor, sensory, visual, and tactile feedback, will help induce neuroplasticity [[ref]](https://link.springer.com/article/10.1007/s13311-016-0423-9)[[ref]](https://pubmed.ncbi.nlm.nih.gov/30508935/).

## 1.2.2 Telerehabilitation

Telerehabilitation can be a highly beneficial option for stroke patients who are willing to receive remote care. For patients with limited mobility or remote access to hospitals, telemedicine can provide a convenient and effective way for patients to receive this care without having to leave their homes. Several research had investigated the quality and patients’ adherence during the telerehabilitation. Just like the traditional treatment, the efficacy of rehabilitation decreases within 6 months, it is recommended that distance rehabilitation exercises be easy, time-consuming, adaptable, personalized or motivating [[ref]](https://www.clinicalkey.com/" \l "!/content/playContent/1-s2.0-S1877065722000951?returnurl=null&referrer=null)[[ref]](https://www.liebertpub.com/doi/10.1089/tmj.2011.0150). One systematic research also affirms the effectiveness of real-time telerehabilitation methods (where medical staff communicate with patients in real time using telephones or live streaming) in alleviating patient discomfort and improving limb mobility [[ref]](https://pubmed.ncbi.nlm.nih.gov/27141087/). In plenty of the rehab methods, MT for phantom limb pain (PLP) patient is confirmed effective to be using in home, and can be conducted entirely with initiation, feedback, and follow-up with health-care professionals remotely [[ref]](https://www.liebertpub.com/doi/full/10.1089/tmj.2015.0009).

On method-based classification, a systematic review broadly categorized telemedicine into three types, and confirms that the telerehabilitation can produce the similar outcome to traditional rehab, it also affirms the importance of cost-effectiveness in telerahabilitation [[ref]](https://pubmed.ncbi.nlm.nih.gov/17697506/):

1. Image-based telerehabilitation:

Using computer vision to quantify patient movements to assess rehabilitation performance, with low network bandwidth requirements [[ref]](https://pubmed.ncbi.nlm.nih.gov/18560142/).

1. Sensor-based telerehabilitation:

Using miniaturized accelerometers and gyroscopes to assess a patient's position and trajectory in 3D space is not a recent development, but few studies have combined the technology with real-time remote communication [[ref]](https://pubmed.ncbi.nlm.nih.gov/16445260/)[[ref]](https://ieeexplore.ieee.org/document/1616295).

1. Virtual environments and virtual reality telerehabilitation:

The definition of the VR telerehabilitation here is “fully immersive virtual environment”, with the use of head mounted visual displays and other sensory input devices like haptic feedback device [[ref]](https://pubmed.ncbi.nlm.nih.gov/15971970/), the system allows a therapist in a remote location to conduct treatment sessions using a virtual-environment-based motor-training system with a patient who is located at home. Compared to the traditional therapy, the advantages of VR therapy can let the patients experience more immersive multisensory feedback, entertainment that improves patients’ adherence and willingness during the rehab.

The above literature surveys indicate that the main axis of research on telerehabilitation is to use communication equipment to try to bring traditional rehabilitation methods into home practice; and to develop small communicable (usually Bluetooth) sensors that can track motion trajectories. Because of its simplicity and relatively good efficacy, the MT-based rehabilitation is capable for investing in routine home rehabilitation. VR rehabilitation is a relatively new research topic. Computer calculations are used to create a realistic rehabilitation environment. In addition to providing a higher sense of immersion in rehabilitation, the digital system is also capable of recording the user's performance. However, the high cost of equipment makes it difficult to put it into home rehabilitation applications.

## 1.2.3 Digital Rehabilitation

In digital rehabilitation, the development of virtual reality (VR) or augmented reality (AR) has received considerable attention [[ref]](https://pubmed.ncbi.nlm.nih.gov/26158918/), various advantages had been found, these including giving patients the motivation of active self-learning, measuring patients’ behavior in safe and realistic environment, and have the ability to formulate more personalized training for patients based on these data, such as dynamically increasing or decreasing task difficulty [[ref]](https://pubmed.ncbi.nlm.nih.gov/30681034/). Several studies have also reported that the VR/AR participants enjoy those rehabilitation session, which can increase their motivation to exercise more frequently at home or between sessions [[ref]](https://pubmed.ncbi.nlm.nih.gov/16586185/)[[ref]](https://pubmed.ncbi.nlm.nih.gov/15502277/). In addition to academic research, there are also many commercial cases in the current VR rehabilitation system [[ref]](https://pubmed.ncbi.nlm.nih.gov/31651335/).

Consider to the cost on the VR/AR equipment and the consumption level in remote regions or developing countries, some of the research topic focusing on a low-cost digital therapy system, dedicate promote the relatively immersive rehabilitation solution. Those research will be discussed in subsection **1.2.5**.

## 1.2.4 The combination of Virtual/Augmented Reality

Based on the above-mentioned advantages, VR/AR technology is often used in combination with other proven effective therapies, and different studies have demonstrated the characteristics of its rehabilitation system: Miclaus et al. using MIRA (non-immersive VR software) to evaluate its effectiveness of upper limb paralysis, The experimental results of 55 stroke subjects showed that the curative effect of MIRA in the six months before and after the stroke was better than that of traditional rehabilitation therapy [[ref]](https://pubmed.ncbi.nlm.nih.gov/32967160/)[[ref]](https://www.mirarehab.com/); Weber et al. designed a 10-subject experiment to examine the feasibility of immersive virtual reality mirror therapy for upper limb paresis after stroke, confirmed that the combination of immersive VR and MT is well-tolerated by chronic stroke patients and has preliminary evidence of efficacy [[ref]](https://journals.lww.com/ajpmr/Abstract/2019/09000/Immersive_Virtual_Reality_Mirror_Therapy_for_Upper.7.aspx); CW Lin et al. developed an immersive virtual reality mirror therapy (VRMT) to evaluate the upper limb function of 30 healthy subjects and 45 stroke patients after using the system for a period of nine weeks. The result was that VRMT had a better curative effect than traditional MT intervention [[ref]](https://ieeexplore.ieee.org/document/9319666)[[ref]](https://journals.sagepub.com/doi/pdf/10.1177/15459683221081430?casa_token=9Aal88GvhVQAAAAA:B79le6WayPglrEBEXr-Y1-T12BDpk3IVGo1Hh9lueiumf5IOBhePBXLveJ_2ctujKwkksGAEaYDL), this study was also the longest and the second-highest number of participants compared to other immersion therapies in a systematic literature review [[ref]](https://www.mdpi.com/1424-8220/22/24/9962); Gilda et al. utilized Electromyography (EMG) to detect the paralyzed arm moving intention from stroke patients, combining the AR technology, rendering the patient's paralyzed arm to the expected position and project it in front of the patient's eyes to provide visual stimulation, found that the range of arm abduction in 4 patients improved significantly after the intervention [[ref]](https://www.tandfonline.com/doi/full/10.3109/17483107.2014.979330?src=recsys). This method of using EMG to predict the movement intention of the user's affected hand and inputting commands into VR has also been used in the other literature [[ref]](https://jneuroengrehab.biomedcentral.com/counter/pdf/10.1186/s12984-021-00822-6.pdf).

## 1.2.5 Immersive Rehabilitation via Mobile Phone

Another research also aims to solve the problem that the cost of VR equipment is too high for telemedicine or personal use. A study shows that the existing VR rehabilitation systems have some limitations, making it difficult for ordinary patients to use them easily, such as relying on a pre-established static safety environment and high costs that make it difficult for individuals to afford [[ref]](https://pubmed.ncbi.nlm.nih.gov/29156493/).

Generally speaking, an immersive rehabilitation system requires at least two sets of independent hardware devices, namely a head-mounted display and a high-performance computer. If we add sensors that capture position, such as the Oculus rift, at least one piece of hardware is confined to a static space, making it difficult to move the entire system, this will lead to be difficult to install and customize to meet specific user needs. As such, most VR rehabilitation systems are limited to hospitals or clinics and are only available to patients during appointments [[ref]](https://pubmed.ncbi.nlm.nih.gov/29156493/). Cost is also one of the key points to consider, without considering the price of a commercial rehabilitation system, the price of additional VR equipment and computing hosts will cost at least 1,000 to 15,000 USD. For hospitals, the hardware acquisition cost of commercial and customized VR solutions may be affordable, because they can be used by multiple patients to share the cost, but for individuals, the purchase of VR rehabilitation equipment is more expensive. The costs may outweigh the benefits.

To solve this problem, Zirbel et al. develop a low-cost, portable, flexible, and interactive VR system called VRehab system [[ref]](https://ieeexplore.ieee.org/abstract/document/8601885?casa_token=yLxZb7LK_w4AAAAA:4RJWtggkz_Bm2VbWtxcpUtRneaOBZghwS02AOLCTAy4BX1DXTTF0BCCWaH9ECtGQO6WLJOE). The system uses a smartphone as a display and computing device, which uses Google Cardboard, a low-cost headset made of lens and corrugated cardboard, and Myo armband worn on the user's arm [[ref]](https://medium.com/thalmic/the-evolution-of-the-myo-armband-24d7e6793f7f). Myo armband predicts hand gestures through EMG signals and uses low-cost Bluetooth transmits the signal to the mobile phone, and then assists the user to make a series of decisions in the virtual environment. The whole thing costs just 100 to 300 USD and can be set up in 10 minutes; Based on the needs of the general stroke rehabilitation, LaPiana et al. has also developed an AR upper limb rehabilitation system using a mobile phone as a platform [[ref]](https://rehab.jmir.org/2020/2/e17822?utm_source=TrendMD&utm_medium=cpc&utm_campaign=JMIR_TrendMD_1), The system uses ARTag, a position marker system that can be applied in AR technology [[ref]](https://ieeexplore.ieee.org/abstract/document/1467495?casa_token=W_es1Cn__j0AAAAA:Hy4U-jsWTe6YcL4Cbk-8h6B4g1bRDEBB9pFCU5FRLrBfi-I8GCSpNkEW-WgGUzIzM6jGMFMU), to locate the position of the patient's hand to complete a series of entertaining tasks.

The lack of in-depth efficacy verification in both documents suggests that research in this field is still developing and has yet to reach maturity.

## 1.2.6 Quantitative Analysis on Mirror Therapy

In the current method of observing neuroplasticity, studies have found that brain-derived neurotrophic factor (BDNF) plays an important role in the mechanism of neurorestoration, and can be used as one of the quantitative indicators for judging rehabilitation status [[ref]](https://www.mdpi.com/2076-3425/10/9/623)[[ref].](https://www.embopress.org/doi/abs/10.1002/j.1460-2075.1982.tb01207.x) Compared to the measurement of BDNF level in the blood sample, another measurement techniques are non-invasive and had been mass adopted by the research. Functional near-infrared spectroscopy (fNIRS) is a non-invasive brain imaging technique that measures changes in blood oxygenation in response to neural activity [ref]. Human tissues are relatively transparent to the near-infrared light (650–1000 nm), it can penetrate the scalp and skull and reach the brain tissue, where it is absorbed by hemoglobin in the blood vessels, by measuring the changes in the absorption of this light, fNIRS can provide an indirect measure of neural activity in different brain regions. fNIRS has several advantages, such as portable, low-cost and relatively easy to access, therefore, some research related to evaluation the effectiveness of MT choose to use fNIRS as their quantitative analysis method [ref][[ref]](https://jneuroengrehab.biomedcentral.com/articles/10.1186/s12984-022-00988-7). Electroencephalography (EEG) is another non-invasive neuroimaging technique, by measuring the electrical activity of the brain using electrodes placed on the scalp, EEG machine can amplify and recorded the activity of certain brain region. EEG’s advantages are high temporal resolution (ability to measure changes in brain activity over time) and its relatively low cost. However, its spatial resolution (ability to precisely locate activity within the brain) is lower than other techniques, same as fNIRS, the MT research also consider using EEG as the analysis method [[ref]](https://pubmed.ncbi.nlm.nih.gov/29853839/). Functional magnetic resonance imaging (fMRI) is a neuroimaging technique that uses magnetic fields and radio waves to measure changes in blood flow and oxygenation in the brain. It provides a non-invasive and safe way to study brain activity. Though it has a best resolution of region and has ability to scan those deep functional cortical area, its enormous size, high cost, and low temporal resolution lead to its low usage in MT rehab research [ref], especially in combination with digital therapy interventions, since fMRI does not allow electronics to enter its interior.

## 1.3 OBJECTIVE

## 1.3.1 Objective of This Study

This study aims to investigate upper extremity rehabilitation in patients with post-stroke hemiplegia. As post-stroke rehabilitation is crucial for patients' recovery, this introduction highlights the significance of neuroplasticity as a rehabilitation theory and recommends evidence-based therapies such as FES, MT, and VR therapy to ensure high-quality care. Furthermore, the COVID-19 pandemic and the rapid growth of technology have led to a growing interest in digital health and telemedicine, offering new opportunities in stroke rehabilitation. However, there are still unmet needs in clinical practice that must be addressed to fully leverage the potential benefits of digital health in stroke rehabilitation.

Stroke rehabilitation typically involves long-term exercises, which can be costly and challenging for patients to perform without assistance from medical staff. To address these issues, this study will explore strategies for increasing the frequency of routine rehabilitation exercises while minimizing the cost of mobilizing medical staff. The primary focus will be on developing a home-based rehabilitation plan that is affordable to the needs of post-stroke patients.

Due to its simple concept and readily available rehabilitation materials, MT has emerged as one of the promising rehabilitation modalities for home rehabilitation during the COVID-19 pandemic. The use of virtual reality technology has also been incorporated into MT to enhance the immersive experience during the practice. Some studies have found that VR-based MT is more effective in improving patients' mobility compared to traditional MT [[ref]](https://ieeexplore.ieee.org/document/9319666)[[ref]](https://pubmed.ncbi.nlm.nih.gov/30681034/). However, expensive equipment costs and the complicated setting process of VR equipment make it difficult to be effectively used in clinical fields [[ref]](https://www.businessinsider.com/why-is-virtual-reality-so-expensive-2016-9" \l ":~:text=A%20high%2Dend%20headset%20like,high%2Dend%20virtual%20reality%20experience.). According to statistical research, the predict of home penetration of VR equipment is only 6.3% in 2026 [[ref]](https://omdia.tech.informa.com/pr/2021-dec/omdia-research-reveals-12m-consumer-vr-headsets-sold-in-2021-with-content-spend-exceeding-2bn).

In addition, how to evaluate the rehabilitation status of home rehabilitation patients will also become a problem that medical staff needs to think about. In the traditional hand-function evaluation scale, the evaluator is often required to be face-to-face with the subject before the evaluation can be performed. By recording the user, The development of digital health will bring the potential of remote assessment.

Affected by some studies that use smartphones, which are more common than VR equipment, as the carrier of immersive stroke rehabilitation systems, we decided to develop a mobile phone augmented real-time rehabilitation system based on the principle of mirror therapy, which is suitable for stroke patients in perform daily rehabilitation exercises at home with their own mobile phone and lightweight, inexpensive equipment.

# Chapter 2 Methodology and Material

## 2.1 SYSTEM DEVELOPMENT

## 2.1.1 Augmented Reality Mirror Therapy System (ARMT)

To solve the defects of the virtual reality mirror therapy equipment mentioned above, this study has developed a mobile device-based Augmented Reality Mirror Therapy (ARMT) system. All data processing of this system is done on a single mobile phone, but users are expected to wear a Cardboard to use (optional). This development aims to allow patients to receive immersive rehabilitation treatments only by using their own mobile device as a display instead of VR head-mounted display. The preview of ARMT is like a general camera preview but allows user to switch between camera mode or headset mode **(Figure 1)**. Under headset mode, the real-time preview screen will be split into two equivalent windows, which are specially provided for the left and right eyes of the human body, resulting in the illusion of three-dimensional feedback. Following the concept of traditional MT, the feature of this system is rendering the image of any human body to the mirror side of the screen, including but not limited to hands and arms covered with clothes. Because of referencing the user's body part as a rendering target directly, ARMT can render different and personalized illusion for each user: the texture of the skin, size, and other characteristics of the mirror fake limb would be rarely different compared to the user's real limb, whereas the background of the preview will not be influenced, this allows ARMT to provide users with more realistic visual feedback than MT, thereby increasing their immersive experience. An additional feature of the ARMT is capable of creating a custom virtual bulletin board object in the AR preview scene for users to notice any necessary information during the rehabilitation.

**Figure 1. Preview of ARMT**

**The preview of camera mode (left) and headset mode (middle). Wearing preview combined with cardboard (right). The left hand in all figures is illusory.**

Based on the concept of augmented reality, not like an immersive VR rehabilitation system, ARMT can adapt to various scenarios without regional restrictions, or the need to define a safe zone, because the user can still perceive the actual surrounding objects and conditions. Considering the cost, versatility, usage rate, and volume of the whole VR equipment, mobile phone-based rehabilitation systems also raise the usability of ARMT, and are suitable for practicing in daily rehab training at any time and anywhere. ARMT's screen update rate can reach a maximum of 60 frames per second (FPS) through internal hardware acceleration. Compared with VR devices, it may not be a smooth level, but it is enough to handle most of the circumstances.

## 2.1.2 Hardware and Accessories

ARMT is only available deployed on iPhones which in iOS 15 (or higher) and equipped with Apple A12 Bionic chip (or other higher end chips). **Table 1** shows the results so far ARMT is applicable to all iPhone models.

**Table 1. All models that can be equipped with ARMT**

|  |  |  |
| --- | --- | --- |
| Launch Year | Models | Estimate Price (USD) |
| 2018 | iPhone XS series, iPhone XR | 230 ~ 260 |
| 2019 | \*iPhone 11 series | 280 ~ 420 |
| 2020 | iPhone SE (v2), iPhone 12 series | 210 ~ 610 |
| 2021 | iPhone 13 series | 460 ~ 860 |
| 2022 | iPhone SE (v3), iPhone 14 series | 310 ~ 900 |

\*Models used for development in this study

In terms of cost, assuming that the user does not own an mobile phone that matches the model of ARMT, the rehabilitation cost of using ARMT is about 300 to 800 USD. Compare to other VR rehabilitation system, even when selecting hardware in the most economical manner possible, the custom-built system will still cost approximately 1500 to 2000 USD [[ref]](https://ieeexplore.ieee.org/abstract/document/8601885?casa_token=yLxZb7LK_w4AAAAA:4RJWtggkz_Bm2VbWtxcpUtRneaOBZghwS02AOLCTAy4BX1DXTTF0BCCWaH9ECtGQO6WLJOE). A Cardboard that meets the iPhone size is recommended to be used together with ARMT rehabilitation. Depending on the material or other additional accessories, the cost of the Cardboard falls between 3 to 20 USD.

## 2.1.3 System Architecture

In order to implement the application that renders the illusion mirror limb based on the appearance of target limb to the real time preview, a efficient trained neural network model for human body segmentation was incorporated into the ARMT system. The neural network model receives a fixed size of pixel buffer from image data stream captured by the smartphone camera, then tagged the set of pixels that are recognized as a part of the human body. Part of the calculation will be accelerated by the Graphics Processing Unit (GPU) and bionic chip, so-called Neural-network Processing Unit (NPU) equipped inside the smartphone, therefore the time consumption from input frame to output result is quite low.

**Figure 2** shows the system architecture flowchart of the ARMT system. The system consist of three components: smartphone, Cardboard and the user itself. The user is responsible for performing any exercise in the field of view of the smartphone back camera, during the system activated period, the camera would continuously capture and transmit real world image frames to the ARMT system. The semantic segmentation algorithm inside the ARMT system will continue to receive these images and complete the calculation in a very short time. If there are any pixels in the frames are classified as belonging to the human body, they will be marked and transformed to mirror coordinates relative to the center line of the screen, and added to the hardware-accelerated shader for rendering to the Graphics User Interface (GUI). In headset mode, the renderer will split the image into two preview images on the left and right, and perform barrel distortion to match the parallax of the worn Cardboard. Finally, the lens on the Cardboard will restore the real time deformed image and provide immersive mirror visual feedback to the user's eyes.



**Figure 2. ARMT System Architecture**

ARMT system is developed by Swift, which is a programming language primarily used to develop native Apple applications, this makes the system extremely adaptable to any Apple devices that meet the specifications and has outstanding computing performance, but it is difficult to transplant to those smartphones with other operating systems.

## 2.2 DEVELOPMENT PROGRESS

## 2.2.1 Development Tools

Developed by Apple Inc, Xcode is an integrated development environment (IDE) that used for developing ARMT. It is specifically designed for macOS, iOS, watchOS, and tvOS app development, mainly supported programming language Swift and Objective-C. Xcode provides a wide range of tools, libraries, and frameworks to facilitate the creation of high-quality applications for Apple devices. Like any other IDE, Xcode includes a code editor that supports syntax highlighting and code completion using IntelliSense.

In actual ARMT development, Xcode allows connecting physical devices and installing the compiled code on the device. It also can monitors some performance indicators of the device when running the program, such as CPU usage, GPU usage, memory space, various thread state, FPS and power consumption estimation.

ARMT has gone through three stages of development, and its characteristics and related frameworks or algorithms will be introduced in the following chapters.

For mobile applications developed without using game engines (such as Unity, Unreal Engine), the User Interfaces (UI) framework will be the starting point of development and the backbone of the code. SwiftUI is a framework provided by Swift, it allows developers to design and develop UI with less code and in a declarative way. In the case of using Xcode, IED supports interface builder to help users directly drag the required basic elements, such as buttons and scrolls, into the view, and don't need to write any code to make. The SwiftUI framework was launched in 2019, and it is recommended that developers gradually shift their development focus to it to replace the old UI design framework UIKit. Therefore, ARMT's overall UI development will use SwiftUI.

## 2.2.2 Hand Joints Skeleton Approach

In AR development, locate the object position in the real-world coordinates is crucial before rendering. Therefore, the main objective to the start of the ARMT system is to capture the actual three-dimensional hand position. For this approach, this study has implemented it based on the combination of 2D hand joints detection algorithm and double vision disparity depth detection. Two main frameworks for developing are provided by Swift: The Vision framework and the AVFundation framework.

The Vision framework is a powerful framework provided by Apple for performing various computer vision tasks on iOS devices. It integrates some built-in machine learning algorithm that utilizes the GPU and neural engine for fast and efficient processing of image and video data. The optimization processes images and video frames in real time, making it suitable for applications that require live analysis and interaction with the camera feed. For 2D hand joints detection algorithm, sourced from the camera feed, this study utilized a hand pose tracking Application Programming Interface (API) provided by Vision to return the two-dimensional coordinates of hand joints and maps these coordinates back to the normalized screen coordinates. This algorithm is learned based on a deep learning model and can be applied in most environments. **Figure 3** shows all 21 hand joint points that can be detected by the algorithm. Algorithms can additionally generate confidence scores from 0 to 1 for detected joints, in ARMT, only the joint points whose current confidence score is greater than 0.7 will be used.

一張含有 文字, 螢幕擷取畫面, 設計, 手 的圖片

自動產生的描述

**Figure 3. Twenty-one hand Landmarks in Vision Framework**

Being widely exist in the organisms on Earth, binocular vision is often accompanied by fusion of vision, although the scene captured by two eyes is similar, the pupillary distance result in a slight difference between the object we saw. This disparity of two images is called parallax, which is one of the stereo vision principles for organisms. Based on the stereoscopic effect produced by binocular parallax, we have the ability to estimate the depth from the camera to the target object by the same scene captured by two cameras mounted parallel. The AVFundation framework is responsible for operating and setting up the camera. In some models of iPhone, there are more than two camera lenses enables double vision disparity depth detection implementation possibility.

一張含有 螢幕擷取畫面, 黑色, 黑暗 的圖片

自動產生的描述

**Figure 4. Ideal pinhole camera model**

**Figure 4** shows the explainable model of dual parallel stereo cameras **C** and **Cʹ**, **Z** is the distance between the target object **P** in the real world and the camera. The distance **B** between the cameras is called the baseline, and **ƒ** is the focal length of the two cameras. Under the premise of fixed baseline and focal length, by triangulation, if we know the pixel points at which positions the target objects are mapped to for the individual two cameras, that is, **x** and **xʹ**,the following equation will hold:

We can know the distance between the target and the cameras **Z**, which is also called depth. For the convenience of description, is called disparity. **Figure 5** intuitively describe the inverse relationship between disparity and depth. When the distance of the target is getting farther, the resulting disparity will be become smaller until it is difficult to estimate, and theoretically it will reach the limit when the disparity is less then one pixel for the lower resolution camera. Likewise, objects should not be too close to the camera, or it would cause over estimation.

一張含有 螢幕擷取畫面, 文字, 字型, 設計 的圖片

自動產生的描述

**Figure 5. The relationship between distance Z and disparity**

According to the description in the iOS developer manual [[ref]](https://developer.apple.com/documentation/avfoundation/avdepthdata), A disparity map provided in the AVFundation framework describes normalized shift values for use in comparing two images. The value for each pixel in the disparity map is in units of 1/meters. Using the actual iPhone 11 for verification, it is found that the depth map generated by it can stably and correctly represent the actual distance between the object and the camera at a depth of 20 cm to 80 cm. This range is enough to be used in ARMT's first-person view hand coordinate detection **(Figure 6)**.

一張含有 室內, 牆, 螢幕擷取畫面, 人員 的圖片

自動產生的描述 一張含有 螢幕擷取畫面, 多媒體, 視訊, 多媒體軟體 的圖片

自動產生的描述

**Figure 6. Depth detection resulting image of camera preview (left) and depth map preview (right).**

**Figure 7** shows the full diagram of hand joints skeleton approach of the ARMT development progress. With multiple back cameras iPhone model, ARMT has the ability to capture the three-dimensional space coordinates of the hand in real time. using the multi-tread execution method to ensure that the hand detection algorithm based on the neural network model will not drag down the screen update rate.

一張含有 文字, 螢幕擷取畫面, 軟體 的圖片

自動產生的描述

**Figure 7. System diagram for ARMT 3D hand joints reconstruction approach**

## 2.2.3 Hand Contour Approach

Several problems aren’t solved in the first approach of the ARMT system development. Including enhance the realistic mirror hand visual feedback like traditional MT, lack of rendering a skeleton and the most serious defect: unreliable 3D predicted coordinates. **Figure 8** shows an inevitable object occlusion can easily make thumb or other fingers’ 3D reconstruction difficult in depth detection algorithms that can only be computed from a single direction of sight In order to avoid reducing the portability of the ARMT system.

一張含有 文字, 手套, 顯示 的圖片

自動產生的描述

**Figure 8. Defects of ARMT depth estimation**

For these purpose, the following development of ARMT system has focused on trying to render a indistinguishable fake mirror hand into the 2D GUI rather than reconstruct the whole hand model into the real world. The users’ hand in the camera feed is the perfect material as a reference. According to the concept of traditional MT, the healthy hand of the patient will be the basis for driving the affected hand. If there is an algorithm that can distinguish the outline of the hand from the background in the image, it will be able to reproduce the realistic affected hand to the mirror side of the preview screen without affecting the presentation of the background like traditional MT.

The study introduced the semantic segmentation neural network model to achieve the above requirements. Semantic segmentation is a field of computer vision that aims to solve the image segmentation problem [ref][ref], especially the background/foreground classification. There are several common techniques that be used to implement in practice, such as Ostu algorithm [ref] and balanced histogram thresholding [ref], these algorithms are based on the image histogram in the classic computer vision image processing. In ARMT’s requirement, the real time hand contour segmentation practice can be seen as a particular case in background/foreground image segmentation. However, it is hard to simply solve it by introducing these segmentation algorithms based on fixed condition judgment. Patients’ skin texture, the color similarity of the hand to the tabletop, and various changing backgrounds of the real world during the rehabilitation training would dramatically influence the stability of such algorithms. On the other hand, the rapid growth of deep learning and the rise of the hardware computing level provides another solution to the image segmentation problem. In several implementation methods of image segmentation, semantic segmentation is a supervised deep learning algorithm to train the model to perceive one or multiple classes of objects in an image, aiming to classify each object at the pixel-wise level **(Figure 9)**.

一張含有 狗, 狗飼養, 螢幕擷取畫面, 草 的圖片

自動產生的描述

**Figure 9. Common applications of deep learning in machine vision**

To train a hand contour segmentation model for the ARMT system, for the training material, this study selected EgoHands dataset [ref], several reasons remaining:

1. The dataset selects 48 different scenario that enriches background diversity, such as indoor office, yard and outdoor table etc. This will help to train the model for all possible rehabilitation environments as much as possible.
2. Totally 4800 image frames labeled pixels of hands appeared in the scene in detail, which is the recorded film of two-person interaction (observer and someone in front of the observer). Basically, each of the images consists of at most four hands: both observer’s and someone’s right and left hand, pixels belonging to each of the hands can be chosen to select or ignored by the developer. These give the trained model potential to semantically distinguish between the observer’s hands and someone’s hands, as well as left and right hands.
3. Using the first-person perspective (observer perspective) to film the images, which is suit for the use of ARMT’s scenario. Because by default, the camera of the ARMT system should be consistent with the user's position and line of sight.

It's worth noting that EgoHands dataset doesn't include gestures, but any outlines of movements that might be made by the hand. In 48 different scenarios, the observer and someone are enjoying some activity like playing cards, chess, and puzzles without being instructed to constrain their hands in a certain gesture **(Figure 10)**.

一張含有 室內遊戲和運動, 圖板遊戲, 桌遊, 遊戲 的圖片

自動產生的描述

**Figure 10. EgoHands image source containing hand contour masks (marked in green)**

The semantic segmentation neural network model deploys on ARMT system need to balance high accuracy and response speed in limited hardware equipment, that is, generate corresponding hand contours in a timely manner based on the image frames provided by the camera. As a lightweight network model, Unet was proposed in 2015 and was originally used to perform fast medical image segmentation at the cell level [ref], because of its simple architecture and efficient performance, it has now become one of the classic semantic segmentation models. On a modern GPU, Unet has the ability to complete the segmentation of several images in one second. For this reason, the study introduces a modified Unet structure neural network attempting to approach the requirement, which backbone introduces the concept of MobileNetV2 [ref]: inverted residuals, linear bottlenecks and depth-wise convolution **(Figure 11)**. According to theory, Unet using this Convolution Neural Network (CNN) structure inspired by MobileNetV2 can reduce the amount of calculation by about 8 to 9 times while maintaining the segmentation accuracy compared to the conventional CNN if both of their CNN kernels are 3x3.

一張含有 文字, 圖表, 螢幕擷取畫面, 設計 的圖片

自動產生的描述

**Figure 11. Modified Unet architectures which inspired by MobileNetV2**

Using transfer learning technique, we applied the MobileNetV2 model and pre-trained parameters in the timm library [ref] to reconstruct the CNN part of Unet to boost the training progress. For the structure of the model is follow the Unet design, 4 down/up sampling in the encoder and decoder part respectively, and those output 4 sets of convolution feature maps number are 96, 32, 24 and 16. Finally, the input dimension of the model is a 224 x 224 x 3 RGB image, which is also the highest resolution that MobileNetV2 can set, and the output is a 112 x 112 x 3 RGB image for the hand contour mask **(Figure 12)**.

一張含有 文字, 螢幕擷取畫面 的圖片

自動產生的描述

**Figure 12. Whole structure of modified Unet with MobileNetV2 backbone**

The process of training the model shows in **Figure 13**, it has been divided into three steps. The first step is data preprocessing. In each epoch, 4,800 EgoHands images are randomly split into a training set of 4,000 images and a verification set of 800 images, and these images will pass random image filter with rotation, horizontal flip, gamma correction, and shifting pixel value based on RGB and HSV color spaces.

一張含有 文字, 螢幕擷取畫面, 字型, 設計 的圖片

自動產生的描述

**Figure 13. Deep learning training workflow**

The second step is to train the model, repeat the training for 200 epochs, and update the parameters in the model by calculating the loss between the source image and the ground truth. The loss function formula is as follows:



Where is dice loss, X is the output result and Y is the corresponding ground truth, The total of dice loss will be between 0 and 1, 0 means that the output result and the ground truth are completely coincident, and 1 means that the two have no intersection at all **(Figure 14)**.

一張含有 文字, 螢幕擷取畫面, 人員 的圖片

自動產生的描述

**Figure 14. The training result of the modified Unet model, the minimum dice loss occurs at the 140rd epoch ()**

The final step of the process is transforming the trained hand contour segmentation model into the format that can be used by the iOS system. CoreML and CreateML are the frameworks provided by Apple that can deliver fast performance on Apple devices with easy integration. CoreML provides several built-in deep learning models for different task including vision, Natural Language Processing (NLP) and sound recombination. CreateML, on the other hand, allowing developers to use third party library to train their custom deep learning model for other applications by using CoreML Converters. As a general-purpose neural network archive format packaged for iOS, the “.mlmodel” file interface can easily preview the outcomes of the model and understand its performance right in Xcode **(Figure 15)**.

一張含有 文字, 螢幕擷取畫面 的圖片

自動產生的描述

**Figure 15. iOS Core ML model conversion workflow**

The concept of rendering related hand contour to the mirror side of the iPhone screen is masking **(Figure 16)**. In SwiftUI design, GUI preview consists of multiple layers of view, these views are sequentially stacked on the GUI and can be masked by their upper view layer and only be shown on the region where the upper view is transparent. After the conversion of CoreML converter, each of the output pixel types from the trained hand contour segmentation model contains four bytes that represent its RGB color value, and the last byte represents transparency. By modifying these values, image frames that only extract the pixel of users’ hand are generated.

一張含有 螢幕擷取畫面, 文字, 手指, 手 的圖片

自動產生的描述

**Figure 16. Multiple layers defined in GUI**

**Figure 17** shows the full diagram of hand contour approach of the ARMT development progress. Since this algorithm does not rely on the depth map, it is also suitable for those iPhones that do not have multiple cameras.

一張含有 文字, 螢幕擷取畫面, 軟體 的圖片

自動產生的描述

**Figure 17. System diagram for ARMT Hand Contour Approach**

## 2.2.4 Cardboard and Screen Fluency Approach

The hand contour approach has greatly improved the fineness of the mirrored hand in the screen preview, but there are still some problems to be solved in practice: Although a neural network model suitable for building on the mobile terminal is selected, it still takes a lot of time to acquire image frames and generate masks; In terms of masking hands, the multi-layer graph stacking method was originally applied to a polygonal mask that does not change over time, and the memory is dynamically allocated to occupy the mask one by one to meet the needs of optimizing rendering efficiency. Forcibly using irregular contours that change every frame will cause frequent dynamic memory allocation and free in a short period of time, which will have a lot of computing costs. There is a strict sequential relationship between the two, and it is difficult to use parallel computing to reduce time costs, which makes the screen update rate of ARMT insufficient. In addition, compared with VR head-mounted displays, ARMT still lacks a good sense of immersion **(Figure 18)**, and cannot be placed on Cardboard for use. It is necessary to rendering the preview that provides the left and right eyes view simultaneously to solve the current problem.



**Figure 18. Rehabilitation of the ARMT system without Cardboard is far inferior to that of VR head-mounted helmets**

To solve the problem of poor performance of the custom neural network on mobile phones, we cite the official human body segmentation model provided by Apple to replace the existing hand contour segmentation model. ARKit is a framework that provides various API for AR experience, including Real-world environment perception, plane detection, creation of virtual object anchors, and label recognition. ARKit3 released in 2019 and added support the human body segmentation API for rendering the occlusion relationship between characters and AR objects **(Figure 19)**.



**Figure 19. The figure occlusion indication shown at the ARKit3 conference**

Although it is hard to peek at the internal algorithm implementation, it is certain that deep learning technology is also used. The algorithm is very likely to adopt the DeepLabV3 [ref] architecture with MobileNetV2 as the backbone of CNN [ref], and optimize it according to the internal computing chip architecture. According to the official statement, the method provided by this API produces a very short masking time, which can make sure the program runs smoothly at 60 FPS. This statement has also been confirmed by actual tests **(Table 2)**.

**Table 2. Comparison of custom hand segmentation model and official human body segmentation model**

|  |  |  |
| --- | --- | --- |
| **iPhone 11** | Custom ML | Official ML |
| Operation time | 35 ~ 40 ms | 10 ~ 20 ms |
| FPS | 25 ~ 28 | 50 ~ 100 |

On the issue of how to render the dual-view preview screen to support the use of the ARMT system on the Cardboard, we quoted the third-party package ARHeadsetKit, which is an open-source project that encapsulates ARKit and efficient rendering methods [ref]. ARHeadsetKit uses Metal frameworks and Single Instruction Multiple Data (SIMD) API for efficient rendering. Metal is the framework provided by Apple to developers to directly access the GPU and can schedule and manage the rendering pipeline according to the usage situation. Metal's memory debugger can maximize the coordination of the workload of the CPU and GPU on rendering graphics, greatly reducing their idle time.

In the rendering task, Metal can manage a piece of memory called frame buffer for the CPU and GPU to share, split the workload required to render one frame into multiple draw calls, and allocate two hardware according to the amount of calculation. In Metal development, shader and texture refer to the programmable stages and image data in the rendering pipeline, respectively. Shader is used to define the various stages in the rendering pipeline, such as vertex shader and fragment shader, which are responsible for processing geometry and rasterization. Texture is used to store image data, which can be used as the input of shader to provide texture, color and other image information for rendering. Compared with the multi-layer stack rendering provided by SwiftUI, Metal tends to use shaders to perform a series of rendering processing for a single texture and push the completed material to the frame buffer for display, this can reduce large amount of dynamic memory allocation and free. In addition to coordinating the workload between hardware, the rendering efficiency is greatly improved.

## 2.3 EXPERIMENTAL DESIGN

## 2.3.1 Purpose

To identify the difference in treatment effects between ARMT and traditional mirror therapy, this pilot experiment is designed to investigate the difference in immediate effects on upper limb performance of the hands between receiving ARMT and MT intervention among healthy subjects. Since MT already had many literatures and experiments to prove its effectiveness in stroke rehabilitation, the ARMT, as the innovative rehabilitation method inspired by MT, needs to show the corresponding evidence in practical circumstances. The whole process of the experiment was held by the therapist from department of rehabilitation in NCKU hospital and the thesis author.

## 2.3.2 Participant Criteria

30 subjects (17 males, 13 females) with the mean age of 23.0 ± 2.7 years old were recruited. randomly divided into two groups and be asked to undergo two one-session interventions, **Table 3** recorded other reference information of clinical subjects before receiving interventional experiments. The subjects recruited in this experiment should have no current or past history of neuromuscular or orthopedic problems in the upper extremities. Both groups received 30 minutes of ARMT or MT rehabilitation, after the one-week wash-out period, changing intervention groups respectively **(Figure 20)**. Before receiving the intervention, all of the subjects must be evaluated their upper limb functionality by the therapists as a pre-test.

**Table 3. Additional information for clinical subjects**

|  |  |  |  |
| --- | --- | --- | --- |
| Ages | Biological sex | Dominant hand | MT experience |
| 24 | Male | Right | No |
| 24 | Male | Right | No |
| 28 | Female | Right | No |
| 20 | Male | Left | Only Once |
| 22 | Female | Right | No |
| 22 | Female | Left | Only Once |
| 19 | Female | Right | Only Once |
| 20 | Female | Right | No |
| 26 | Male | Right | No |
| 20 | Male | Right | No |
| 22 | Male | Right | No |
| 23 | Male | Left | Only Once |
| 27 | Male | Right | No |
| 20 | Female | Right | No |
| 20 | Female | Right | No |
| 19 | Female | Right | No |
| 23 | Female | Right | No |
| 25 | Male | Right | No |
| 28 | Female | Right | No |
| 23 | Male | Right | No |
| 28 | Male | Left | No |
| 22 | Male | Right | No |
| 21 | Male | Right | No |
| 26 | Female | Right | No |
| 25 | Male | Right | No |
| 20 | Male | Right | No |
| 20 | Male | Right | No |
| 27 | Female | Right | Only Once |
| 23 | Male | Right | No |
| 22 | Female | Right | No |

一張含有 文字, 螢幕擷取畫面, 圖表, 行 的圖片

自動產生的描述

**Figure 20. Flow chart of the clinical trial in healthy subjects**

## 2.3.3 Procedure

The MT condition (with a mirror box) or ARMT condition (including an iPhone and a VRG headset) were deployed in the experiment **(Figure)**. For all healthy subjects, they were told to sit on a chair and placed both of their hands on the table, and their right hand was instructed to stay still. This is for the purpose of regarding right hand side as an affectation of stroke patients’ hemiplegia side.

During the first 10 minutes of intervention, NIRScout (fNIRS, NIRx Medical Technologies, Glen Head, NY, USA) equipment was used to collect the changes in oxyhemoglobin in specific brain cortex regions of the subject, with a total of 16 detectors and 8 sources, arranged in an array 3 cm apart and distributed over the upper scalp of the motor cortex and prefrontal cortex **(Figure)**.The position refers to the channel position established by other literature that use fNIRS as MT efficacy analysis. For MT experiments, hand movement must be involved, so almost all literatures will choose the motor cortex on both sides. In addition, some literatures want to observe the sense of coordination and immersion of MT subjects through fNIRS or fMRI. For this purpose, the prefrontal cortex is usually selected as the observation Region of Interest (ROI) [ref][ref][ref][ref]. The experiments in both MT and ARMT conditions were arranged in a 1-minute block design repeated 10 times, including the first 30 seconds rest stage with eyes closed and the last 30 seconds of finger pinching trials, an audio clue indicates the subjects to open and close their hand, perform 2 times of pinches in 1 second **(Figure 3)(Figure)**.



**Figure 3. Block design of the experiment for each condition**

In the last 20 minutes of intervention, subjects will be instructed by a professional therapist to continue performing repetitive movements for hand rehabilitation while using MT or ARMT (Table 4)(Figure). After the whole 30-minute intervention experiment, the subjects will immediately receive the hand function assessment post-test.

## 2.3.4 Hand Function Assessment Tool

For the outcome measurements, several standardized upper limb tests are used to be pre- post-test assessment tools such as Pinch-Holding-Up-Activity (PHUA), Purdue Pegboard Test (PPT), Semmes-Weinstein monofilament (SWM), Minnesota Manual Dexterity Test (MMDT). These assessment criteria will be assessed by a professional occupational therapist. **Table 5** organized the type of functions each standardized upper limb test represents, Each of these will be described below.

**Table 5. The content of the upper limb tests**

|  |  |  |
| --- | --- | --- |
| **Abbreviation** | **Category** | **Index** |
| PHUA | Hand strength coordination | 1. Maximum pinch force 2. Ratio of max pinch force and load force |
| PPT | Finger dexterity | 1. Pins insertion count 2. Pins assemble count |
| SWM | Finger sensitivity | 1, Touch-pressure threshold |
| MMDT | Upper extremities gross motor function | 1. Time spent on placement 2. Time spent on flipping |

PHUA test is one of the tools for evaluate human hands coordination in sensory-motor coupling controll. In most of the circumstances, humans’ strength of pinching or grasping an object would passively adjust from the changing weight of current object causing by unstable weight center or acceleration, trying to found a best efficient ratio between strength and the holding object load weight. Based on this behaviour, Flanagan et al. had proposed the ratio of the maximum grip strength value to the maximum load value during the lifting process can be regarded as a sensitive parameter by which the generation of grip force relative to the load due to momentum can be observed [ref]. The standard pinch-holding equipment of PHUA is the metal cuboid weighing 480 grams, which contains two load cells and two accelerometers to measure subject pinch strength and detect the cuboid acceleration when it moves **(Figure)**. In the beginning of the PHUA evaluation stage, the subject will follow the instruction from the therapist to lift the metal cuboid above the desk for 5 cm and make sure only holding with their thumb and index finger on load cells in the entire process. After 5 second, according to the computer sound prompt, the subject raised the cuboid to 30 cm. Data collection takes 15 seconds in one epoch. The subjects performed a total of 3 times, resting for 1 minute each time **(Figure)**. The parameters analyzed include (1) maximum grip force value (FPPeak): the maximum pinch force (pinch force) during the ascent of the device (from 5 to 30 cm); (2) maximum load value (FLMax): at the maximum The maximum load force during acceleration; (3) force ratio: the ratio of FPPeak to FLMax. The lower this value (the ratio is closer to 1), the better the subject has better force control.

PPT is a fine fingertip dexterity test with validity and reliability. Subjects were asked to arrange pins and assemble pins with washers and collars within the limit of time. Results of unilateral (dominant and non-dominant hand) pin insertion count (30 seconds), bilateral pin insertion count (30 seconds) and numbers of assembly (1 minute) consisted in outcomes (Figure). The higher the score, the more pins completed within the limited time, which is used as a reference to express the finger dexterity of the subject.

SWM uses Touch Test Sensory Evaluators (North Coast Medical, Inc, Gilroy, CA) to test the skin pressure threshold. Nylon filaments of different thicknesses are applied to the volar skin for 1-1.5 seconds of continuous force, and the threshold is defined as three tests. The lightest stimulus with at least two correct answers. The test will be applied to the pulp of the finger. Each filament is numbered as the logarithm to the base ten of the strength of a tenth of a milligram. The smaller the value, the higher the sensory sensitivity of the subject's fingers **(Figure)**.

MMDT is a test that measures gross motor function of the upper extremities. The equipment consists of a number of round cake-shaped building blocks and two wooden boards with grooves for placing the building blocks. In this study, the subjects were asked to stand in front of the MMDT equipment to perform placing and flipping all the building blocks, and they were not allowed to move their positions again during the process **(Figure)**. The evaluating therapist would record the time it took to complete the placing of the building blocks and flipping all the building blocks. The less time spent, the better the subject performed on gross motor skill of upper extremity.

## 2.3.5 Brain Area Activity Measurement

At the same time, for the sake of obtaining objective quantitative data, this research additionally collects the changes in cerebral blood perfusion to analyze bilateral motor cortex and prefrontal cortex activation via fNIRS when intervention begins. In terms of quantitative data, uses Homer3 [ref] to remove signal artifacts, a bandpass filter (0.01~0.1 Hz) to remove physiological signals that may affect (heart rate, respiratory, etc.), and finally output the changes in oxyhemoglobin after block average in the specific cortex **(Figure)**. The General Linear Model (GLM) of the modified gamma function convoluted square wave according to the time length of the intervention interval was also applied for effect estimation, according to assumptions, the beta value generated by GLM indicates the strength of the modulation of the hemodynamic response [ref].This value can be used as indirect evidence for the activation of the relevant brain area (Figure).

# Chapter 3 Result and Analysis Conclusion

## 3.1 RESULT AND COMPARISION

## 3.1.1 Hand Function Evaluation

**Table 1** shows the difference in outcome measures between pre- and post-training in each assessment scale for 7 healthy subjects (recruit 10, 3 haven’t attended the trial) within the mean age of 22.9±2.8. Change in **FRPeak** and percentage of maximal pinch strength show statistically significant differences within ARMT condition (*p=.018* and *p=.028*). Also, a significant effect was detected in **PPTDH** (*p=.027*), **PPTBH** (*p=.026*), **MMDTplacing** (*p=.018*), and **MMDTturing** (*p=.018*). By contrast, none of the outcome results show significant differences in MT group. However, not have a significant difference between ARMT and MT groups, either. The observed insignificant differences between the groups may be attributed to inadequate sample size, or it implict that the ARMT exhibits comparable effectiveness to the MT when evaluated against identical testing criteria.

**Table 1. The outcome measures in each assessment scale.**

**\*DH: dominant hand, \*BH: both hands.**

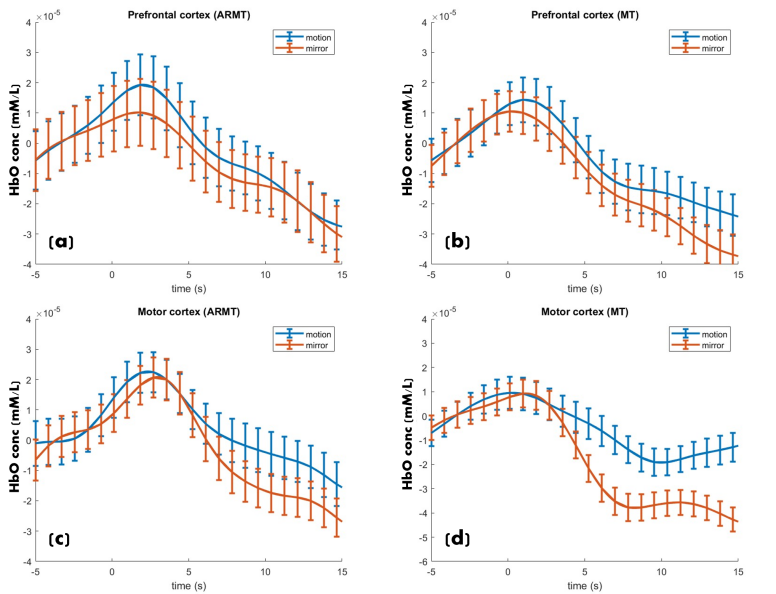
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Effectiveness**  **(INDEX)** | **PRE** | **ARMT** | | **MT** | | **Group Difference** | |
| **POST** | **P Value** | **POST** | **P Value** | **Z Score** | **P Value** |
| **PHUA**  **(FRPeak)** | 3.55±.59 | 2.93±.41 | **.018\*** | 3.26±.52 | .249 | -1.352 | .176 |
| **PHUA**  **(Percentage)** | 0.42±.13 | 0.32±.08 | **.028\*** | 0.41±.16 | 1.000 | -1.016 | .310 |
| **PPT**  **(\*DH)** | 14.86±1.57 | 17.07±1.34 | **.027\*** | 16.00±1.26 | .125 | -1.581 | .114 |
| **PPT**  **(Non-DH)** | 14.93±1.40 | 16.00±1.41 | .090 | 15.29±1.91 | .672 | -.631 | .528 |
| **PPT**  **(\*BH)** | 12.43±1.21 | 13.43±1.30 | **.026\*** | 13.43±1.17 | .206 | -.085 | .933 |
| **PPT**  **(Assembly)** | 41.57±3.85 | 45.00±4.10 | **.018\*** | 44.50±3.21 | .271 | -.169 | .866 |
| **SWM**  **(Thumb)** | 2.39±.04 | 2.39±.04 | .317 | 2.39±.04 | .655 | .000 | 1.00 |
| **SWM**  **(Little Finger)** | 2.38±.04 | 2.37±.03 | .317 | 2.41±.04 | .157 | -1.732 | .083 |
| **MMDT**  **(Placing)** | 64.67±7.82 | 58.36±5.64 | **.018\*** | 60.16±7.82 | .176 | -1.69 | .866 |
| MMDT  (Turning) | 46.35±5.20 | 38.71±3.65 | .018\* | 41.04±5.29 | .128 | -.676 | .499 |

## 3.1.2 fNIRS Result of ROI

9 healthy subjects (4 males, 5 females) participated in this trial and contributed a well-identifiable fNIRS signal. For channel-wise comparison, the ipsilateral hemisphere of the subject’s moving hand (left hand) will be called the mirror side (left hemisphere), while the contralateral hemisphere will be called the motion side (right hemisphere).

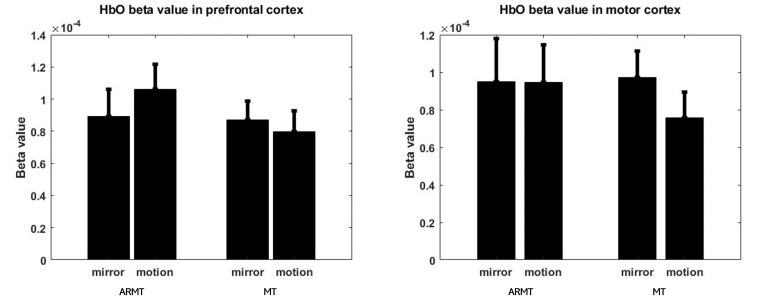
In the subjects’ prefrontal cortex region, both MT and ARMT group observed a tendency towards blood perfusion during the intervention **(Figure 4)**, it takes about 3 seconds to reach its highest peak after the start of the intervention and then quickly drops below the baseline. Besides, greater activation was observed on the motion side.

In the motor cortex region, longer-lasting and more pronounced activation was observed in the ARMT group, And the peak of the mirror side is slightly later than the motion side. Although the blood perfusion of HbO activation at the mirror side was not greater than that of the motion side in any circumstances, a highly correlated pattern of signals was still revealed.



**Figure 4. Trends in HbO during the intervention. (a) prefrontal cortex in ARMT group, (b) prefrontal cortex in MT group, (c) motor cortex in ARMT group, (d) motor cortex in MT group.**

**Figure 5** shows the motion side of ARMT group has the highest activation at the prefrontal cortex in the mean of HbO’s beta value of GLM. But the subjects who received MT have higher activation levels on the mirror side. This phenomenon is also shown in the motor cortex area. However, there were no statistically significant differences between groups or bilaterally (*p>.05*).



**Figure 5. Mean beta values and corresponding standard errors at the prefrontal (left) and motor (right) cortex.**

## 3.2 RESESRCH PERSPECTIVE

## 3.2.1 Discussion

## 3.2.2 Conclusion

# Chapter 4 Epilogue

## 4.1 ARMT Potential on Stroke Rehabilitation

This study indicates the potential of the ARMT system in the field of rehabilitation at home. Using their personal mobile device, patients in need have an alternative method to traditional MT without requiring additional equipment. The ARMT system can provide a more realistic view that enhances the immersive experience. However, it remains to be seen whether the effectiveness of rehabilitation using ARMT is greater than that of traditional MT, in the purpose of approaching the answer, more clinical trial subjects should be recruited.

## 4.2 Limitation

The limitation of this pilot study is the small number of subjects, which may cause individual differences to have a significant impact on the trial's results. Another limitation of the study is that the healthy subjects included did not have a history of stroke or any brain damage. Therefore, their blood perfusion signal response in the certain cortex may not be representative of the general behavior of stroke patients. Subjects with a history of stroke or hemiplegia will be considered to participate in the experiment in the future.

## 4.3 Future Work

1. \* Student [↑](#footnote-ref-0)
2. \*\* Advisor [↑](#footnote-ref-1)