Master in Music Sound and Computing
Universitat Pompeu Fabra

Digital music interfaces for motor rehabilitation: a motion capture and machine learning approach

Raquel Lucena Peris

Supervisor: Rafael Ramírez Meléndez

August 2023

Universitat Pompeu Fabra Barcelona

Master in Music Sound and Computing

Universitat Pompeu Fabra

Digital music interfaces for motor rehabilitation: a motion capture and machine learning approach

Raquel Lucena Peris

Supervisor: Rafael Ramírez Meléndez

August 2023

Contents

1 Introduction **1**

1.1 Motivation 1

1.2 Objectives 2

1.3 Research question 3

1.3.1 Is the combination of music and artificial intelligence (AI) useful for stroke rehabilitation? 3

1.3.2 Is the artificial intelligence digital music interface comparable to a musical instrument while doing stroke rehabilitation? 3

2 State of the art **5**

2.1 Introduction 5

2.1.1 Stroke 5

2.1.2 Music Therapy 6

2.1.3 Music and Stroke 7

2.1.4 Music, Technology and Stroke 8

2.2 Related Work 9

2.2.1 Objectives and direction of study 10

2.2.2 Area of rehabilitation 11

2.2.3 Data acquisition 11

2.2.4 Classification algorithms 12

2.2.5 Feedback 13

2.3 Goals and open work for this project 13

3 Methods **15**

3.1 Materials . . . 15

3.1.1 Hardware . . . 15

3.1.2 Software: interface and implementation . . . 16

3.1.3 Pipeline . . . 19

3.1.4 Algorithms . . . 20

3.2 Design of experiment . . . 21

3.2.1 Goal experiment . . . 21

3.2.2 Current experiment . . . 22

3.3 Sessions structure and evaluation . . . 23

3.3.1 Stroke patients' sessions . . . 23

3.3.2 Without stroke patients' sessions . . . 24

3.4 Sessions evaluation . . . 24

3.4.1 Music Therapist Evaluation . . . 24

3.4.2 Patients Evaluation . . . 25

3.4.3 Healthy Participants Evaluation . . . 26

3.5 Interface development . . . 27

4 Results **32**

4.1 Algorithms comparison . . . 32

4.2 Quantitative evaluation . . . 34

4.3 Qualitative Evaluation . . . 35

4.3.1 Music therapists evaluation . . . 35

4.3.2 Patients evaluation . . . 36

4.3.3 Healthy patients evaluation . . . 37

5 Discussion **39**

5.1 Digital music interface . . . 39

5.2 Limitations . . . 40

5.3 Future work 41
5.3.1 Complete the study 41
5.3.2 Improve the cognitive and musical game 42
5.3.3 Increase versatility 43
5.3.4 Avoid delay 43

6 Conclusion **44**

List of Figures **46**

List of Tables **47**

Bibliography **48**

A GitHub Repository **51**

Acknowledgement

First, I would like to express my sincere gratitude to Dr. Rafael Ramírez, my thesis supervisor. His advice, instruction, competence and support have been essential to advance at every step of the thesis.

I also give thanks to all the music therapy group that was present during the hospital sessions. A special mention goes to María and Gonzalo, the music therapy master's students that collaborated with me on the hospital sessions under the guidance of Paolo. Their insights and feedback refined and enriched my work.

The thesis could not have been done without the participation on some tests and the support of patients, family, friends and other master's students. This project is especially dedicated to each of you.

Finally, thanks be to God, who made it all possible. *Whoever remains in me and I in him will bear much fruit, because without me you can do nothing. Jn 15:5.*

Abstract

In this work, a digital music interface software has been developed for helping motor rehabilitation in stroke patients by capturing the movement using a webcam.

To improve patients' quality of life using the digital interface, the Processing programming language has been used to design the interactive interface. Some Python scripts have been used to load the deep learning models that detects the body keypoints. Finally, the Wekinator software has been used to learn the positions and predict poses from new incoming data poses.

As a result of the previous steps, the interface offers different tasks to improve motion on patients. Two of the tasks have been designed to be used simultaneously with a music therapist or a musical song. On the one hand, while the patient's position is mapped to a musical sound, the music therapist accompanies the patient with an instrument. On the other hand, a song is selected with the interface and reproduced. The patient keeps the tempo with the movements and accompanies the song with some music notes or drums. The other two tasks are designed so that the patient can use them without much help; it is a game that combines cognitive and motor rehabilitation.

The interface has been tested on a few stroke patients and on some healthy people to test its potential. After the sessions, some questionnaires were done to: evaluate the results and compare the digital instrument with a traditional instrument, asses the understability and utility of the interface and check the progress of the patients. The results show that the digital instrument interface has potential as a tool for stroke rehabilitation, but a longer and larger study needs to be done to have robust results.

Keywords: Machine Learning; Interactive interface; Pose estimation; MediaPipe; MoveNet; Motor rehabilitation; Stroke; Music Therapy.

Chapter 1

Introduction

Since the time of Ancient Greece, the idea of using music to improve health has existed. After the First and Second World Wars, music was used to reduce physical and emotional trauma. From that moment on, the need to prepare and train musicians to be able to complement medical therapies was evident. In the 1940s, music therapy was developed from a professional clinical point of view for the first time, some people involved in pioneering music therapy were: Ira Atshuler a psychiatrist and music therapist that applied music therapy on Michigan, Willem van de Wall who used music therapy using funds of the Unitate States, and Thayer Gaston who played a decisive role in the advancement of the profession from an organizational and education point of view. [1] In this paper, a digital music interface is developed to allow the use of professional clinical music therapy in stroke patients.

1.1 Motivation

This project aims to give an answer to the physical limitations that stroke patients suffer in their movement. According to the World Health Organization, annually, 15 million people suffer a stroke [2]. The number of strokes is increasing due to the aging population, as the most important risk factor is age. For stroke survivors, there will always be sequels that will impact their quality of life. To counteract this loss in quality of life, a digital music interface is proposed as a gamification technique to motivate and commit people to rehabilitation exercises using music therapy. Music can be an excellent tool to improve physical rehabilitation by increasing motivation on physical therapy, as it can offer a way to improve mood and express feelings or emotions. Acording to *Sociedad Española de Enfermería Intensiva y Unidades Coronarias* [3], music therapy reveals a reduction in anxiety, stress and pain. As a traditional instrument is difficult to play with physical disabilities, the digital interface can be a good tool to implement to try to reduce stroke effects.

1.2 Objectives

The digital music interface that is going to be described should achieve the following objectives to create a useful new product for stroke rehabilitation:

- **Works in real-time.** If the interface gives an immediate response to the patients using it, the experience will be more motivating and engaging for users. In the ideal case, a patient's movement will trigger a sound at the same time, with no delay between the gesture performed and the generated sound.
- **Is a personalized interface.** As stroke affects each individual in several different ways, the rehabilitation needs to be unique to each patient. For this reason, the digital interface needs to be personalized with different options, parameters or settings that could be selected by a physiotherapist, a music therapist or the patient to solve the variety of limitations that exist after a stroke.
- **Learns in real-time the user's capabilities by machine learning (ML).** A ML algorithm implemented on the interface will allow adaptability to learn the rehabilitation positions of patients. If the system can learn and analyze the patient's movements during the process, the sound response will take into account the evolving capabilities of patients, reducing lack of motivation and abandonment of rehabilitation.
- **Has a validation method.** To ensure that our system can really be applied to the rehabilitation of stroke patients, a good validation method is needed.

It will be necessary and important to see the evolution (improvement or deterioration) in the movement capacity of the patient in the rehabilitated area. These validations can be done through the annotations of the medical professionals after the implementation of some sessions (in short and long term) and also through the feedback of the music therapists and patients after having taken part in some of the rehabilitation sessions.

1.3 Research question

In this section, two research questions are described to try to guide the investigation towards the utility of the digital music interface used for motor rehabilitation in stroke patients:

1.3.1 Is the combination of music and artificial intelligence (AI) useful for stroke rehabilitation?

With the idea of using music and AI in stroke rehabilitation, the utility, profitability and advantages of using the interface will be analyzed by evaluating and studying the results obtained. It is expected that the use of music will be engaging for the patients; for this reason, the rehabilitation sessions would be more beneficial, proving the utility of music on stroke rehabilitation. Besides, if AI is helpful for motor movements' improvement, it will be another advantage to use it in rehabilitation.

1.3.2 Is the artificial intelligence digital music interface comparable to a musical instrument while doing stroke rehabilitation?

This question wants to answer the objective that we have to bring the interface as close as possible to a musical instrument. In this way, the patient can feel how, through the movements he is making, one musical result or another is obtained. At the same time, providing patients a digital musical instrument will give them the opportunity to perform exercises independently. It is expected that users will have control over the musical outputs; however, it might be difficult to allow expressiveness of the musical output during the sessions.

Chapter 2

State of the art

2.1 Introduction

In this chapter, the research topics of the thesis are discussed regarding the current work done for and automatic system rehabilitation in stroke patients using a digital music interface.

2.1.1 Stroke

Stroke is the interruption of blood flow to the brain because a blood vessel has been broken or blocked. The blood does not reach an area of the brain; therefore, the affected nerve cells die because they do not receive oxygen. If some brain area receives poor blood flow, its function may be temporarily or permanently changed. There are two types of stroke defined [4]:

- Ischemic: blood flow to the brain is blocked and the brain can not get nutrients and oxygen, causing brain cells to die in minutes. This type of stroke mechanism in involved 90% of the cases.
- Hemorrhagic: sudden bleeding in the brain because of a vessel rupture. The leaked blood pressures the brain cells and the surrounding central nervous system, damaging them.

According to the Hospital Clínic de Barcelona [5], one in six people will suffer a stroke during their lifetime. The Spanish population experiences a stroke every 6 minutes, and a patient dies every 14 minutes. Stroke is the second leading cause of death worldwide; it kills more than 6 million people every year. It is also the main cause of physical disability in adults and the second cause of cognitive impairment.

2.1.2 Music Therapy

Music therapy is a treatment that uses music elements as an intervention to improve motor coordination, neurological function and mood in patients. Music therapy uses music, rhythm and beat to retrain the brain. There are two different types of music therapy, active and passive.

- **Active** music therapy is where the patient directly sings, plays an instrument, or moves with the music during therapy.
- **Passive** music therapy, also known as sensory music therapy, involves patients listening to familiar music.

In terms of music selection, active music therapy chooses music with a strong sense of rhythm, while passive music therapy mainly chooses music based on personal preferences.

Depending on the therapeutic area, Guttmann [6] defines that the music therapy techniques used can be divided into three broad categories:

- **Physical Intervention:** Focuses on techniques for developing strength, balance, movement, and coordination through rhythmic patterns.
- **Stimulating communication:** Using vocal techniques and singing well-known songs can help stimulate speech and improve pronunciation, clarity, and expression, and help patients memorize vocabulary.

- **Cognitive stimulation:** Using a variety of active and passive techniques involving playing and listening to music can help patients improve concentration, memory, and problem-solving skills through the use of specific musical activities.

Participating in music therapy does not require any knowledge of music. The ability of music to stimulate responses makes music therapy an important part of many patients' treatment plans.

2.1.3 Music and Stroke

Music therapy can help stroke survivors restore a variety of functions, including but not limited to motor, language, and cognition. According to FlintRehab, a leading provider of restorative neurorehabilitation to improve the way stroke survivors recover [7] there are several benefits of using music therapy for stroke patients:

1. Promotes improved gait patterns (walking). Music therapy is able to treat affected gait patterns through a technique called 'rhythmic entertainment'. In this technique, the movements are synchronized with the rhythm.

2. Improves the function of the affected hand with strength, range of motion or dexterity (manual skill).

3. It helps to improving the speech abilities of patients with aphasia. Speaking is usually a function regulated by the left brain, while singing is a function regulated by the right brain. The therapist may use singing therapy to teach individuals how to speak again through singing.

4. Promote the recovery of memory, attention and other cognitive functions after stroke.

5. Assists with easing post-stroke depression music treatment as the brain releases more dopamine anticipating climaxes in the music, improving the mood and reducing feelings of sadness and depression after stroke. This is significant, as almost one in every three survivors battles with post-stroke depression.

6. One out of five stroke survivors is affected by post-stroke anxiety. Music therapy can reduce cortisol, the stress hormone.

7. Practice an activity with high repetition helps to stimulate the brain. The more you play out a particular activity, the better the cerebrum gets at executing that capability. If music therapy is combined with engaging exercises, the rehabilitation can be better.

The benefits we are more focused on are related to motor rehabilitation, as mentioned in [8], music therapy utilizes a distributed network of brain regions, including the auditory and primary motor cortex. Therefore, as music therapy increases motor cortex excitability, which is also associated with recovery of motor functions, using music for stroke rehabilitation has a major impact on stroke recovery as the auditory cortex is connected to the motor cortex. The motor system is very sensitive to sound stimulation, and the rhythm of music can stimulate motor neurons to make muscles move naturally. In addition, studies cited in [9] have shown that patients that receive rhythmic auditory stimulation achieve significantly better outcomes on step velocity, stride frequency, symmetry and length, balance function and trunk control.

2.1.4 Music, Technology and Stroke

Since specially trained music therapists are not always available, time spent with a music therapist is valuable and limited. Attending music therapy once a week is not enough for complete rehabilitation.

To stimulate the brain and maximize recovery, it is important to continue music therapy at home. By using technology and music therapy, an efficient path to rehabilitation can be obtained. One way to achieve this is by using rehabilitation techniques that combine music with the areas you want to work on.

2.2 Related Work

In this section, there is a review of previous systems that have been used in the rehabilitation of stroke patients (or other diseases) as gamification rehabilitation systems. These are models proposed that provide some musical or game framework for stimulating motor coordination, plus some creativity and visual or auditory perception. There are a huge variety of products that combine music, technology and rehabilitation on stroke patients. These new systems with a musical framework may be ideal for patients' motivation.

First, a list of tools that combine music and technology for stroke rehabilitation is being explained with a little detail:

1. **Music Glove:** Flint Rehab's MusicGlove [7] uses movements synchronized with music to promote hand function. To use this rehabilitation device, the patient needs to put the glove on the affected hand and perform various pinching movements in synchronization with the music game. This rhythmic movement helps stimulate neuroplasticity and rewires the brain. Furthermore, the average user accomplishes hundreds of repetitions per half-hour session with MusicGlove. All the data recorded by the sensors are saved to a database.

2. **GenVirtual:** A product developed by the Universidade de São Paulo [10] created an augmented reality musical game that stimulates creativity, concentration, memorization, visual and auditory perception and motor coordination. The player needs to use his body to interact with the game in an imaginary situation where the real elements predominate in rhw virtual environment by following a sequence of sounds from a random or known melody with the markers, which are associated with notes. This system has some benefits as it facilitates motor planning by associating markers with notes. Due to the association, patients can make their own corrections with the sound generated. A computer webcam captures images of the upper limb which are processed to identify the markers and the notes associated with them.

3. **GUI for the Rehabilitation of Elderly People:** A musical tangible user interface is proposed by the University of Milan [11] with three different option games. This interface is expected to provide training and physical exercise in a fun and enjoyable way.

4. **Drums PAD:** Among other products that combine music, technology and rehabilitation on stroke, the last one explained in more detail is developed by the University of London [12]. The patients hit the digital drum pads with their hands or objects at the tempo of a self-chosen song. The rehabilitation area of this product was the affected upper limb, and after the sessions, some questionnaires were done to get some data measures.

There are many products and papers that combine music and technology to improve the health of stroke patients or study the topic, not just the ones mentioned. For this reason, in the following subsections some of them are mentioned and different features are obtained for each: the main goal of the paper, the area of the body that is being improved, the way in which the data is obtained and the classification algorithms that are used.

2.2.1 Objectives and direction of study

Some studies have the objective of studying the correlation between tested machine learning or deep learning algorithms for rehabilitation and comparing them with clinical trials. Some of them, as in *Accuracy of machine learning algorithms for the assessment of upper-limb motor impairments in patients with post-stroke hemiparesis* [13] concluded that automated systems can provide good support to therapists. There are some studies as [14], that go further and try to improve these algorithms that have achieved a good accuracy.

Most of the previous research designed a system to do rehabilitation in patients as in: *GenVirtual: An Augmented Reality Musical Game for Cognitive and Motor Rehabilitation*[10], *Counteracting learned non-use in chronic stroke patients with reinforcement-induce movement therapy*[14], *Motivating stroke rehabiliation through* *music: A feasibility study using digital musical instruments in the home*[12] and *A Music Tangible User Interface for the Cognitive and Motor Rehabilitation of Elderly People*[11]. In some of the studied papers, the systems can be applied to other diseases such as Parkinson's disease, spinal cord injury, Huntington's disease, multiple sclerosis and cerebral palsy.[13] Not only diseases, but paper [11] talks about other types of decline in elder people such as reduced processing efficiency, working memory, attention, physical activity, response time.

The idea of using algorithms for improve rehabilitation is not only studied in humans, a research [15] used rodent data to research the deep learning approach to assess stroke impairments in mice.

2.2.2 Area of rehabilitation

In most cases, the studies are centered on studying the rehabilitation of the upper limb (arm) after a stroke as [13], [10], [11] or [14].

2.2.3 Data acquisition

In order to carry out the different studies to facilitate the rehabilitation of patients, it is necessary to obtain some data. In each study, a different number of data and samples is required, which depends on the task to be solved and the algorithm used. The set of data has been collected in different ways for each study. In terms of samples of data, in some studies [13] the experiments result in improvements in performance when the data size increases. In other systems, as described in [15] or [16] the algorithms used for extracting body points are pre-trained with huge datasets as ImageNet and then in the classification part, only few training examples are needed. Furthermore, a study [15] uses principal component analysis to reduce the dimensionality of the data. On the following list, there are the different techniques and devices used for recollecting the data in all the studied works.

- Surface electromyography (sEMG) which measures the electrical signal of the muscle recorded from the skin. [13]

- Electroencephalography (EEG) which measures the electrical activity in the brain using electrodes attached to the scalp. [13]

- Inertial measurement unit (IMU) that is an electrode that measures velocity, orientation and gravitational.[13]

- Cell phone.[13]

- Microsoft Kintect motion capture system [13] [14]

- Sensors, which are a combination of accelerometers for orientation and direction, gyroscopes and magnetometers, track the movements of patients. [7] Also other features as time doing the task as [12] use.

- DeepLabCut and DeeperCut offer user-defined feature tracking using reduced costs from video recordings, a computer vision algorithm that performs automatic and markerless tracking of user-defined features. This idea is on *Deep learning-based behavioral profiling of rodent stroke recovery*[15].

- A computer webcam is used to capture images with a view. Data is captured, analyzed and processed using a deep learning pose estimation algorithm to identify the markers. [10] [16]

- A deep learning pose estimation algorithm is used for capturing the data. This method is followed in another master's thesis *Development of a Virtual Reality Application Focused on Rehabilitation Therapies Combining Virtual Reality glasses and Deep Learning*[16].

2.2.4 Classification algorithms

There have been used different Machine Learning and Deep Learning Models to perform the rehabilitation in patients: Support Vector Machine (SVM), Random Forest (RF), K-nearest neighbor (KNN), Artificial Neural Network, (ANN), Decision Tree, Extreme Learning Machine (ELM), Linear discriminant analysis (LDA), Convolutional Neural Network (CNN) or Backpropagation Neural Network (BPNN). On the paper that studied accuracy in ML algorithms in rehabilitation [13] the strongest correlation between clinical assessments and algorithms was achieved using CNN, DT, SVM and ELM.

2.2.5 Feedback

ome of the systems studied proposed ways for extracting information after the rehabilitation, enabling the rehabilitation program to be adjusted and optimized in function of the results. This data can be used to set individual goals for each patient. Some ideas for having feedback are:

- Saving all the sensors, images and devices information obtained in the performances to a database[11] that could be later analyzed to accurate the tracking process.
- Doing questionnaires after the rehabilitation sessions where participants can report their opinions.[12]

2.3 Goals and open work for this project

In the different subsections of the related work, we could have seen that there are different approaches to do rehabilitation on stroke patients, some of them proposed a gamification system related with music. This thesis aims to develop a musical digital interface for stroke rehabilitation using the webcam of a computer to capture the data using a machine learning approach.

The quality of the final system will depend on several parameters: the characteristics of the patients, the quality of the data recorded defined by the motion capture model chosen, the quality of the camera, the final classification algorithm used, the library of sounds defined and finally the user interface.

According to *A primer on motion capture with deep learning: principles, pitfalls and perspectives* [17] a pose estimation algorithm can be seen as a function that maps each image taken by a webcam as the coordinates of body parts. There are several algorithms that allow human pose estimation that have been trained with a large dataset, some of them are developed in Tensorflow and others by Google:

- **MoveNet** Lightning is the state-of-the-art pose estimation model available in a smaller, faster and less accurate version than Thunder version, which is used for cases that require higher accuracy.
- **MediaPipe** provides a suite of libraries and tools for you to apply AI an ML techniques to the digital interface.

Once a capture model has been defined, the next step is the creation of an user application that allows:

- Defining a number of different rehabilitation classes and the number of photos per class desired. The creation of the dataset is relevant for pose recognition and the size can be small. It has to take into account the resolution of the images taken to extract the keypoints; a lower resolution will occupy less space if images are saved on disk and will allow detecting keypoints faster, resulting in velocity, but the results will be less accurate.
- Training the created pose classes with the TensorFlow or MediaPipe Model to obtain their keypoints and save them to a dataframe.
- Using one of the classification algorithms that performed better in the current task or in previous work and make some predictions. The algorithm can be evaluated from some Python libraries as scikit-learn and tested using the Wekinator Classification Algorithms.
- Using the predictions to generate sounds. These sounds can be either generated with a Processing library or downloaded from a library of sounds.

Chapter 3

Methods

3.1 Materials

The digital music interface allows the user to control: the body area to be rehabilitated, the exercise to be done and the sound that will be generated during a rehabilitation session.

The developed system consists of an interface developed in Processing, an open-source Java-based program language. This interface allows interaction with the user to choose the desired rehabilitation exercises. The user can switch between four exercises; all of them are motor rehabilitation oriented: two of them are more focused on music therapy and the others are centered on cognitive elements.

3.1.1 Hardware

This system does not require a lot of physical materials. The following elements are enough:

- A computer with a webcam capture the movements and the gestures.
- A loudspeaker will help to listen properly to the generated sounds.

- A current computer mouse or a touchscreen will allow interaction with the digital interface.

3.1.2 Software: interface and implementation

The software programs, languages, libraries and models will be the most important part of the materials; they are the main contributors to the digital music interface and implementation.

Python and models

Through the user's interaction with the interface, the area that is going to be rehabilitated is selected. There are three main body parts: face, hands and full body. Different artificial intelligence trained models are used for the rehabilitation task, the models run in Python executable to identify the movement on: right hand, left hand, both hands, face, all the body, body without face and body without face and lower limb. Face and hands landmark pose detection are using MediaPipe, an open-sourced framework developed by Google that allows building machine learning solutions for live media. MediaPipe Face/Hands Landmarker outputs 3-dimensional landmarks in real-time. Then, for body keypoints detection, MoveNet is used. MoveNet is developed by TensorFlow Hub and is an ultra-fast and accurate model that detects 17 keypoints in a body. There are two variants of Movenet: Lightning and Thunder. We are using Lightning as it is faster than Thunder. To use the detection models, some scripts of Python were implemented. The main structure of the Python scripts are:

- Download the required packages.
- Import the downloaded packages.
- Initialized the model.
- Set the port where the keypoints are going to be sent as OSC messages.

- Open the camera, start to process images with the model loaded and send the body points detected via an OSC port.

The Python files (.py) were exported as executable (.exe) files to avoid manually running the scripts. So once a body area is clicked by the user, one of the executables starts to run. Once the Python executable is running, the webcam is taking video frames and the hands, face or body artificial intelligence models are sending the keypoints detected via a Port OSC.

Wekinator and machine learning

Wekinator is the software that listens to the Python outputs and allows to build an interactive machine learning model by human interaction instead of writing programming code. To initialize Wekinator some parameters need to be set: the number of inputs which depends on the body area that is being tracked, the listening port where Wekinator is receiving inputs from Python (this is set by the programmer and the value is an integer) and the output port where Wekinator is sending values to Processing (as receiver port, this is set by the programmer and the value is an integer), the number of outputs (number of models that are being trained on Wekinator, in this task is always 1), the type of Machine Learning algorithm we are using on the classification task and the number of classes are being trained (this value is an integer and depends on each rehabilitation task and patient). There is a text box on the digital music interface to help users set those parameters. In Table 1, the values used to fill the Wekinator parameters are defined.

Table 1: Paramater settings on Wekinator

Receiving OSC	6448 *integer*
Number of Inputs	*Both Hands:* 83. *Left Hand:* 42. *Right Hand:* 42. *Body:* 37. *Torso and Arms:* 16. *Body (no face):* 24. *Face:* 936.
Number of Outputs	1
Output Port	12000 *integer*
Type	All classifiers (default settings)
Options	KNN. AB. DT. SVM. NB.
Number of Classes	2-8 *integer*

Once all the parameters are defined, to create the different body classes, the 'Start

Recording' button allows recording data for each of the classes. Once all the classes have been recorded, the Train button will train all the classes and finally, by pressing the 'Run' button the Machine Learning algorithm starts to do real-time classification with the new data that is coming as input. This new real-time classification is the output of Wekinator that is sent by the output port to the Processing file that is running with the digital music interface.

Processing and sound

Processing software is one of the most commonly used by visual designers and it is the software used on this digital music interface. The interface has all the interactions needed that will allow th personalization of all the system. The Processing software also receives the outputs from Wekinator which are processed in the code and mapped to a musical note or a drum sound.

Library of sounds and images

A collection of drum sounds and songs has been created to be used during the rehabilitation sessions. Also, some images have been downloaded that are used on the interface.

Questionnaires

To obtain some feedback from each session and some thoughts about the interface, three different questionnaires have been created using Google Forms. The first one is addressed to the physiotherapist that is present during the rehabilitation session, the second one to the music therapists that learn to use the program and the last one to users of the program (whether they are patients or not). After a period of sessions, the results of the questionnaires are extracted to a *csv* file and a Python Notebook is used to analyze and extract information from the data.

3.1.3 Pipeline

To sum up the material section, figure 1 and figure 2 shows the steps that the system does to train the classes, and how it works in real-time. First, during training, the interface is opened and the rehabilitation type is selected. Then a body area is selected and the webcam is opened. The webcam is taking body points of the selected body area with the MediaPipe or MoveNet models. The points are sent to Wekinator which creates the class that will feed the machine learning algorithm.

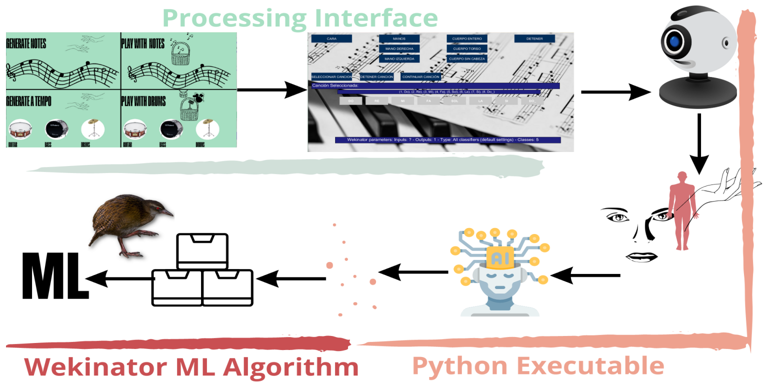

Figure 1: Digital music interface system training pipeline.

During the test part, we are receiving images of the same body part as the training. The models extract the body features which are sent to Wekinator and a new class is predicted. The predicted class is sent to Processing and generates some sound or action.

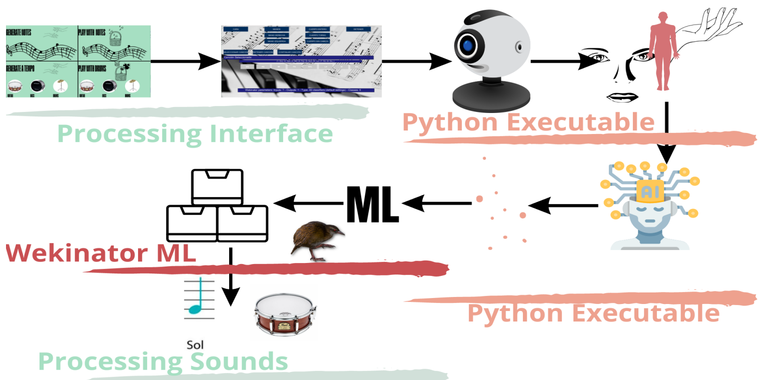

Figure 2: Digital music interface system testing pipeline.

Finally, on figure 3, there is a whole schema of the different programming parts and the inputs and outputs that they are using.

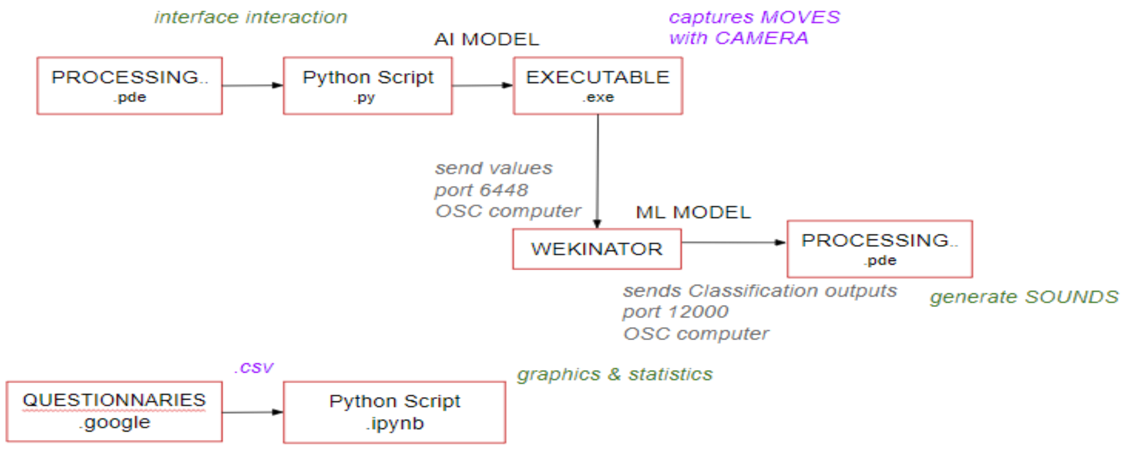

Figure 3: Rehabilitation system workflow.

3.1.4 Algorithms

Different algorithms are applied to classify the inputs (landmarks of the body extracted by the models). Various ML algorithms have been considered to explore and detect the most effective classification algorithm for the interface.

Support Vector Machine

Support vector machine is a supervised learning method.[18] It finds a hyperplane in an N-dimensional space that classifies data points. The benefits of using that algorithm are that it is effective in: high dimensional space and memory and it is also flexible as different kernel functions can be used. There are some weak points: when the number of features is much greater than the number of samples over-fitting will occur.[19]

K-Nearest Neighbour

K Nearest Neighbour is a non-parametric supervised learning method. The algorithm stores all the data and classifies the new data based on similarity measures to the neighbours. K refers to the number of nearest neighbors to include on the similarity measure. The output of the algorithm is the majority class among the new data K-nearest neighbours.[19]

Decision Tree

Decision Tree is a non-parametric supervised learning method. It creates a model that predicts a new target value using decision rules inferred from the data features. The advantages are that it requires little data, is able to solve multi-output problem and the result is easy to interpret. Some disadvantages are that over-fitting is easy if some mechanisms are not taken into account (setting a minimum number of samples of each class or pruning) and small differences in the input data can predict a completely different output value. [20]

Ada Boost

Adaptive Boosting is an ensemble learning technique that first applies a classifier to the data samples and then fits copies of the classifier to the same data, adjusting the weights of the incorrectly classified training outputs. [21]

Naive Bayes

Naive Bayes is a supervised learning algorithm that applies the Bayes' theorem assuming independence between features. It does not require a lot of training data to estimate the parameters. [22]

3.2 Design of experiment

This section explains in a little more detail which is the experiment was designed to be able to carry out a study of the evolution of rehabilitation in patients, and which has actually been carried out.

3.2.1 Goal experiment

In an ideal situation, to carry out the experiment, it is necessary to have a control group that receives rehabilitation sessions without applying music therapy and an experimental group that receives sessions where music therapy and physiotherapy are combined. To carry out a complete experiment, it is also needed to have a minimum number of patients (10 control and 10 experiments) to be able to extract robust and reliable results.

A number of sessions are necessary over a period of time to observe the evolution and effects of the rehabilitation on the patients. Eight sessions in total over a month, two sessions per week, might be enough to apply the digital interface to patients and extract some results.

It is essential to have an objective method to evaluate the sessions. If it is a medical recommendation, a good option is to do the *Fugl-Meyer* test: a viable, useful and easy-to-apply scale, which makes possible to assess the functional status of stroke patients.

3.2.2 Current experiment

At the time of carrying out the real experiment in Centre Fòrum - Parc de Salut Mar, the reality has been that although there have been two months to carry out the experiment, June and July 2023, only very few sessions have been possible to do. The study has been limited to only 2 patients, since during the hours in which my presence was scheduled, there were only 3 patients. The first patient was already part of the study of the final master's project of another internship student, so he was not part of my study, although one day he tried this digital music interface. The professionals at the hospital have preferred that the second patient did not take part in the study because of health issues. Finally, there was a third patient who could participate in four rehabilitation sessions where the digital interface was applied. Although the music therapist and I have been in the hospital waiting for him more days, due to exhaustion, anxiety and mental situation, he decided not to come, following the medical recommendation. This patient who participated in the experiment did not realize movements that were part of the *Fugl-Meyer* test; for this reason, another measurement was applied.

Due to the limited number of patients, the interface was also tested on people who had never suffered a stroke or lost movement. This allows checking the functioning of the interface.

3.3 Sessions structure and evaluation

In this section, a time scheduled for a single session is proposed.

3.3.1 Stroke patients' sessions

Stroke patients int the hospital are meant to spend 30 minutes doing a session, but the reality is that most of the time the sessions last around 20 minutes. There is a slight difference in the first session that a patient receives compared to the following sessions.

First session: On the first session, about 10-15 minutes are used to explain the system and its utility, as well as to record and train the classes for each patient. The next 10-15 minutes are used to start the rehabilitation movements while patients are generating sounds with their bodies. A music therapist accompanies them with a classical instrument. The exercises consist of the repetition of some movements; each one of this movements should last around 30 seconds, as patients tire quickly. After some repetitions of the exercises, the patient needs time to rest, relax and control the respiration. At the end of the session, two different questionnaires need to be filled out: the first one by the music therapist and the second one by the physiotherapist.

Other sessions: On the following sessions, we spend 5 minutes preparing all the system, and if it is needed, another 5 minutes can be spent to add some more new training data to the classes if the patient's movements are a bit different from the previous sessions due to improvements in the movements. New training data will be useful when the patient starts to recover mobility and his movements are different from the previous ones recorded. The rest of the session would be to exercise (with some pauses) the injured body part with music accompaniment. At the end of each of the sessions, only the physiotherapist's form should be completed to obtain some evolutionary data about the patients.

3.3.2 Without stroke patients' sessions

For those who have never suffered a stroke, the session lasts approximately 15 minutes. The initial five minutes are dedicated to explaining what the interface is for and how it works while the programs are being executed. The next 3 minutes are used to test the interface with 2 classes of movements (using drum or note sounds), then the next 3 minutes are used to test the interface with 3 classes using the other kind of sound. Finally, a questionnaire of 2 minutes is given to the users to obtain some feedback.

3.4 Sessions evaluation

In order to evaluate the application, some metrics and results are needed. Currently, the only objective metric that has been considered is the precision of different machine learning algorithms on the data. In order to obtain more information and be able to evaluate the interface created, three different questionnaires have also been created. The results of the questionnaires might be subjective, but we can extract some conclusions from them. In two of the forms, the questions were based on two other articles. First, *A framework for the evaluation of digital music instruments* [23] gives the idea of getting different evaluations from various points of view. Then, following *The EyeHarp: A Gaze-Controlled Digital Musical Instrument* [24] some of the proposed questions have been adapted to the current work.

3.4.1 Music Therapist Evaluation

Link to the questionnaire: https://forms.gle/M3DzZZKczZvxaA3NA

This questionnaire aims to study the understability and utility of the interface. There are different questions and each one has its own purpose in the study.

- *Cause comprehension: were the available input gestures clear?*: to analyze if the music therapist finds it clear and easy to understand all the available movements .

- *Effect comprehension: were the available control parameters clear?*: to evaluate whether the control settings that affect the performance of the interface are easy to adjust.

- *Mapping comprehension: was the connection between the input gestures and the control parameters clear?*: to study if there is a clear connection between the input gestures and the control parameters that is understandable by the music therapist.

- *Intention comprehension: how well did the system allow the user to express his musical intentions?*: to analyze whether the musical interface allows the patients to express their musical purpose or not.

- *Error comprehension: if there had been errors in the performance, would they have been noticeable?*: to study if the performance errors are very noticeable or not if you are using the system.

- *What motivation does sound add compared to not having it?*: to investigate the benefits of having sounds on the system in contrast to not having sounds.

3.4.2 Patients Evaluation

Link to the questionnaire: https://forms.gle/evPshw8daX6dbCfZ6

The goal of this questionnaire is to evaluate the progress of the patient using the *Fugl-Meyer* scale if possible; otherwise, there are open answers to keep track of the progress and to assess motor functioning. This questionnaire is meant to be answered by the physiotherapist if possible; if not, a professional music therapist with knowledge on rehabilitation should answer the questionnaire. On the first session, the patient's diagnosis is used to obtain information on the patient's current health status. Then the rehabilitation area is defined, and when using the *Fugl-Meyer* scale some questions are asked with closed options to obtain answers; otherwise, there is an open answer question to include the observations of the patient's exercise.

3.4.3 Healthy Participants Evaluation

Link to the questionnaire: https://forms.gle/qfmTMmmHeCDYUgUD6

Using the answers of some healthy participants, this questionnaire aims to compare the digital music interface with a traditional music instrument.There are different questions, each one has its own purpose in the study:

- *How much previous practice and training does the performer need for performing with the instrument when compared to a traditional musical instrument?*: to assess how much practice is needed to play on the digital instrument and if it is easier to play than a traditional instrument.
- *How much control does the performer have on the musical output?*: to analyze if the users perceive a great control to express and control the output sounds.
- *How much real-time feedback (e.g., visual or auditory) does the user receive from the system?*: to study the amount of information (feedback) that the user receives from the digital music interface.
- *How tiring is it to play music with this system when compared to musical instruments?*: to check the tiredness of the system in comparison with a traditional musical instrument.
- *Is it hard to play in tempo with the system when compared to musical instruments?*: to find if it is more difficult to play with the digital interface on tempo in comparison to a musical instrument.
- *Which approach between the rhythmic accompaniment and the note accompaniment do you consider more user-friendly?*: to analyze the preferences of the user, the question asks which accompaniment the users want: rhythmic or note accompaniment.

- *What degree of synchronization do you find between gesture and sound?*: to check the degree of synchronization between the input gestures and the output sounds that the users perceive.

3.5 Interface development

The main part of the project is to design a digital music interface that meets the objectives. The most important objectives are that the interface can be personalized for each patient and works in real time. It is also important that the interface can be easily used by the professionals during the sessions.

Throughout the process of developing the digital music interface for the final master's thesis, two interfaces have been proposed and used. Next, the characteristics of the two interfaces are detailed.

First proposal

The first interface developed uses only Python programming language. On the figure 4 there is the interactive digital interface proposed.

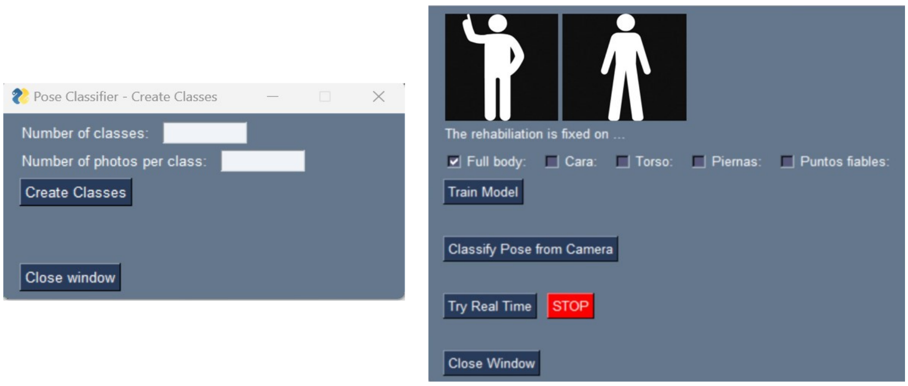

Figure 4: First proposal Python interface

Using the Python interface is straightforward to train the system for users. The user that uses the platform can set the number of classes and the number of photos per class to create each one of the classes used for training. Then, by selecting the checkbox, the user can decide in which part the deep learning model has to extract keypoints. Finally, by just clicking the following buttons: train model, classify pose from camera and try real time, the user can start to use the digital interface on real-time, so that new input gestures are mapped to sounds.

Second proposal

The second proposal was developed to improve the previous interface; the most important correction was avoiding the time delay that existed between the input gesture and the output sound generated.

Figure 5 shows the first display that shows the interface. There are represented four different interactive tasks which can be grouped in 2: two of them are thought to be accompanied by a music therapist or a song, and the other two can be seen as a game to play.

Figure 5: Second proposal Processing interface

The tasks that need to be accompanied by some music is represented on 6. This first task has the objective of rehabilitating the desired part of the body through music therapy. For this reason, when carrying out this task, the presence of a music therapist who accompanies the patient with a traditional musical instrument is optimal. In cases where there is no musician, there is the possibility of selecting a previously downloaded song on the computer; thus, the patient or user will have to follow the tempo of the selected song instead of being accompanied by the musician.

Firstly, the rehabilitation area has to be selected: face, hands or body. When one of these buttons is pressed, the script that contains the Python executable starts to run, so the points of the body are beginning to be detected and sent them in messages through the ports. The Wekinator program that allows the classes to be trained is also opened, the message displayed below on the task interface allows you to configure the parameters.

In addition, it can be observed that there are eight buttons corresponding to the C major scale. By default, class 1 corresponds to the note C, class 2 to the note D, and so on up to 8 classes. In order to generate a better accompaniment, you can change the mapping between classes and musical notes, so if you press any of the buttons it is assigned to the first available class. In this way, if it is decided to press the notes C, E and G, class 1 will correspond to C, class 2 to E and class 3 to G, the other classes will keep the same default class. This new mapping will appear in the message above the note buttons.

Figure 6: Rehabilitation task visualization

Another task follows the same idea as the one defined above, but now a maximum of 3 classes are defined that correspond to three drum sounds. The sound of the drums also need to be accompanied by the music therapist or it is necessary for the user to follow the rhythm of a selected song.

The other task does not need the accompaniment of the music therapist or the selected song. Its goal is not only to improve the movement of stroke patients but also to exercise the cognitive part of patients. On Figure 7, first the therapist has to select the body area to be rehabilitated, train the data using Wekinator and start to try the interface in real-time. In this task, only two classes must be defined; one class will be mapped to generate the figure moving to the left side, and the second class is mapped to move the figure to the right. The game consist of catching the notes that are falling. To catch them, by performing the first or second gesture, the figures are controlled. When the two-person figure touches the musical note, the punctuation increases and the musical note is played. Eeach time that 100 points are gained, a new note appears until 5 different musical notes are caught.

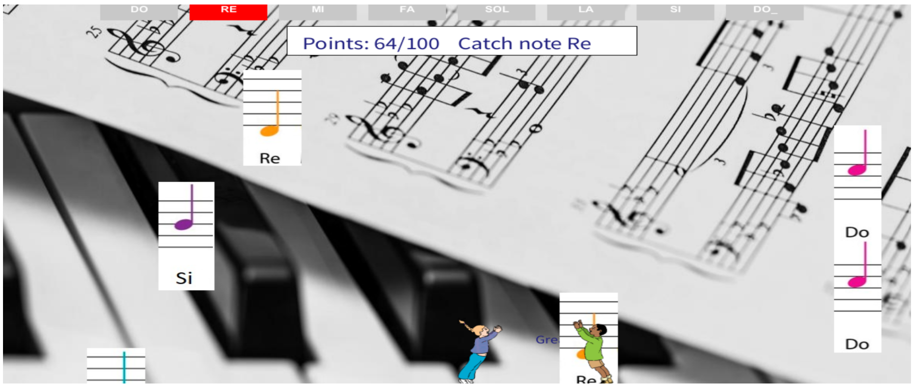

Figure 7: Rehabilitation game task visualization

As before, there is the same game where instead of musical notes, the falling icons are drums, and there are only three kinds of drums.

Proposals comparison

After the development of both proposals, finally the one used in hospital and on healthy patients was the second proposal. In the following Table 2 there are some features that allow a comparison of both proposals.

As stated at the beginning of this section, the most important feature of the interface is that it can work in real time. For this reason, the second proposal was chosen.

Proposal 1	**Proposal 2**
1 programming language (Python)	2 programming languages (Processing & Python) + Wekinator program
Easier to train for users	More tricky to train
Difficult to add new training data	Easier to add new training data
Difficult to code	Easier to code
Simple interface	Better visual designed interface
Huge Delay (0.3s)	Slight Delay (0.05s)
More space needed (all images are saved)	Less data saved

Table 2: Comparison of interface proposals

Some functions of the digital music interface of the second proposal have not been implemented on the first one, for example, the game task. The second interface uses Processing language for creating the interface. This allows having a better visual interactive designed interface which improves the user experience. Moreover, the data that Wekinator uses to train the model is saved as some numerical values and to add new data on another session day it can be easily done by clicking a button. However, on the first proposal the data that is being saved is currently images and the option to use new data without deleting the previous one is not implemented. So another advantage of the second interface is that all the Machine Learning programming can be avoided, which simplifies the coding part and the Wekinator program for training the data does not need to save a large amount of data.

The clear benefit of the first proposal compared to the second one is that the training part is easier. On the second proposal, the professional that assists in the rehabilitation session has to know how to use Wekinator to create the classes and train them.

Chapter 4

Results

In this chapter, the results obtained during the process are presented. Firstly, results on machine learning algorithms accuracy are presented. Secondly, the results obtained from the questionnaires are shown on different graphics and analyzed.

4.1 Algorithms comparison

Different Machine Learning algorithms have been tested in both Python and Wekinator to get the best classification algorithm. Wekinator results are very difficult to analyze and study, since the only thing we get from there is the final classification when the algorithm is tested in real time. It is impossible to obtain objective training or test metrics using Wekinator. For this reason, some Python code, modified from the first proposal, has been used to be able to compare different algorithms and obtain different accuracy values between different classes.

In the next two tables, the precision has been calculated using two different deep learning models. Each row of the table corresponds to the accuracy that has been obtained by comparing the number of classes considered among them. For example, the accuracy value using SVM algorithm between 2 classes is the average value obtained from the training accuracy between classes 2 and 3, class 2 and class 4..., and so on, comparing all classes with all.

In this first table 3, eight classes have been created using full-body movements. The first class was with the body upright and both arms down; the last class is the same position but with the right arm raised. All other classes have an arm position between these two positions.

Table 3: Accuracy of each algorithm using body gestures using different number of classes

Classes considered	SVM	KNN	DecisionTree	AdaBoost	NaiveBayes
2.0	0.678571	0.982143	0.901786	0.919643	0.955357
3.0	0.529762	0.979167	0.952381	0.988095	0.934524
4.0	0.466071	0.969643	0.814286	0.944643	0.9625
5.0	0.339286	1.0	0.910714	0.8875	1.0
6.0	0.285714	0.958333	0.830357	0.645833	0.9375
7.0	0.1875	0.946429	0.767857	0.5	1.0
8.0	0.125	1.0	0.75	0.625	0.9375
All	0.445007	0.978191	0.877434	0.889136	0.962129

In the second table 4, ten classes have been created using hands movements. Each class number equals the same number of fingers up.

Table 4: Accuracy of each algorithm using hand gestures using different number of classes

Classes considered	SVM	KNN	DecisionTree	AdaBoost	NaiveBayes
2.0	0.773333	1.0	0.986667	0.991111	1.0
3.0	0.335417	1.0	0.957292	0.998958	1.0
4.0	0.38	1.0	0.985714	1.0	1.0
5.0	0.236264	1.0	0.96398	0.991758	1.0
6.0	0.166984	1.0	0.936508	0.964444	1.0
7.0	0.39537	1.0	0.915741	0.944907	1.0
8.0	0.186667	1.0	0.922222	0.937778	1.0
9.0	0.195652	1.0	0.93913	0.878261	1.0
10.0	0.36	1.0	0.92	0.68	1.0
All	0.303668	1.0	0.955148	0.979253	1.0

Results show that KNN and NB are the algorithms that work better in both cases. However, the previous tables 3 and 4 show the accuracy obtained using only the training and validation sets. This means that the gestures that are being predicted come from the same dataset that was used for training. Later, a test dataset has been created with the same gestures. The images now are very different: the size is not the same, the location of the person doing the gestures is different and the background has changed. Using the classifier fit to classify 8 classes of body gestures and 10 classes of hand positions, the accuracy table using the test dataset are:

Table 5: Body validation & test accuracy of each algorithm

Dataset	**SVM**	**KNN**	**DecisionTree**	**AdaBoost**	**NaiveBayes**
Test 8 classes	0.25	0.75	0.125	0.25	0.25
Validation 8 classes	0.186667	1.0	0.922222	0.937778	1.0

Table 6: Hands validation & test accuracy of each algorithm

Dataset	**SVM**	**KNN**	**DecisionTree**	**AdaBoost**	**NaiveBayes**
Test 10 classes	0.1	0.1	0.1	0.2	0.1
Validation 10 classes	0.36	1.0	0.92	0.68	1.0

The results are not good; the only value that is quite good for the body gestures test dataset is the accuracy obtained using the KNN algorithm. These poor results may also be due to the fact that the camera is located in a completely different location than the training and validation data. There are changes in the brightness as well as in the angle and orientation of the camera positions. In addition, the distances at which the images between train-validation and test dataset are captured are different, which means that there are changes in scale. It could also be that there are other factors, such as background changes, that affect motion detection. The results obtained with the Python algorithms are poor, despite that, as every patient is going to do the rehabilitation task using the same camera, the same hospital room and more or less the same position it is expected that in real-time using Python the results are better. When using the second proposed system in real-time, the classification is not that bad. Wekinator algorithms can be quite different, but it is not possible to compute some metrics as the accuracy.

4.2 Quantitative evaluation

At the end, no physiotherapist was present during any session and any scale test (as *Fugl-Meyer*) was followed during the sessions. For that reason, all the possible evaluations are quite subjective and qualitative.

4.3 Qualitative Evaluation

4.3.1 Music therapists evaluation

One music therapist and three students answered a questionnaire to study the understability and utility of the interface. Figure 8 shows in green the answers that indicate a good feature of the interface, in yellow the ones that can be improved and in red those that are not well achieved with the interface developed.

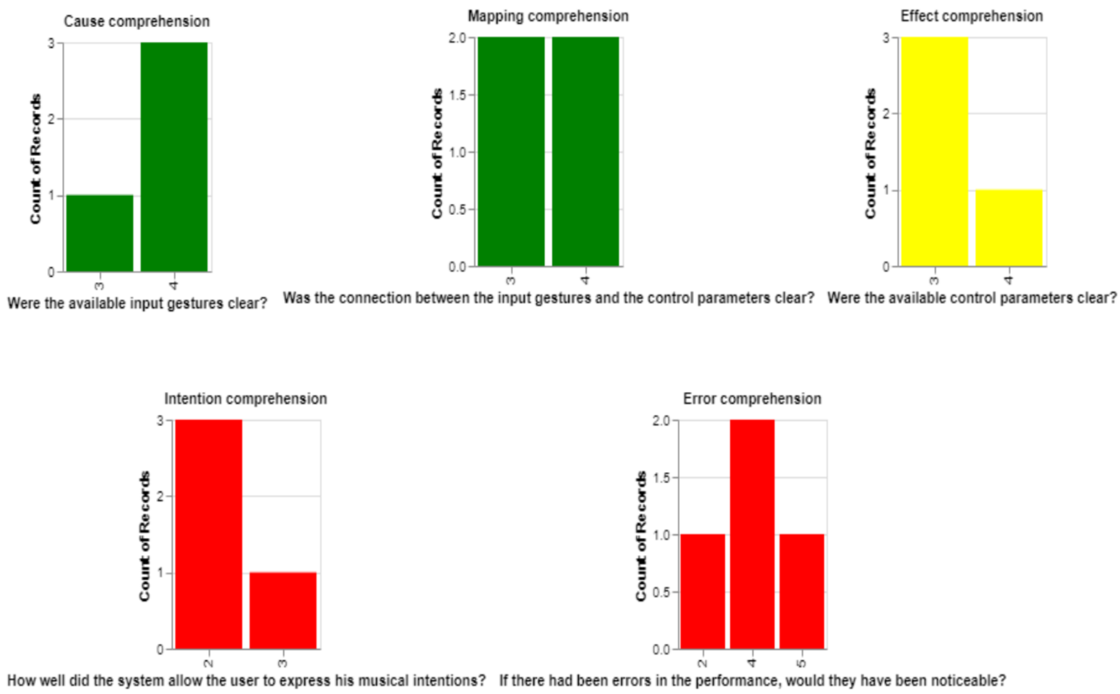

Figure 8: Music Therapists' Evaluation Graphics

All the subplots answer some question and clarify the comprehensibility and the usefulness of the interface. All answers vary in a range from 1 to 5. Analyzing the results, we see that both the gestures that the patient can make for rehabilitation and the relationship it has with the input parameters to be configured are quite clear. Even so, it is not so clear how to configure these parameters. Finally, the music therapists negatively rate the expressiveness that the interface allows and also say that an error is quite noticeable (since the sound is generated).

4.3.2 Patients evaluation

For the evaluation of the patients, it has not been possible to obtain an objective metric; therefore, the information we have after each session has been the comments that the music students have provided. On table 7, for each patient, there are the diagnosis, the rehabilitation and the comments for each session that show the evolution of the patient.

Table 7: Summary of patients' diagnosis and evolution of sessions

Patient	**A**	**B**
Diagnosis	Ictus & anxiety	Ictus
Exercise	Arms and shoulders extension exercise from the position of the center of the chest to the back of the head. The exercise is performed with a rod weighing about 200 grams.	Raise the arm from the rest position to above the head, keeping it straight at all times. The movement is accompanied by a piano scale and the program plays drum sounds.
Sessions	1: Patient A presents pain when it goes down to the neck area approximately. The patient performs the exercise well and can do several repetitions.	1: Patient B performs the exercise well but gets quite tired, and in each full turn you have to stop for a few seconds, and every certain time we do only one piano turn to control breathing. 2: There are improvements in the resistance of the movement, and there is a better synchronization with the piano. 3: The patient controls his movements; sometimes he accelerates and does not follow the tempo. He starts to control the breathing and he seems more calm and comfortable. 4: New data is trained as the patient's movements become wider. The patient has control over his own breathing and the synchronization with the piano is more correct.

Patient B participated in more days of rehabilitation sessions. With it, it was possible to see that during the time the rehabilitation sessions lasted, there was an improvement in some aspects: the patient learned to control his breathing while performing the rehabilitation task, he learned to perform the rehabilitation exercise more autonomously (without having to constantly indicate to him how to do the exercise well) and he also started to listen to the music accompaniment and follow the same rhythm.

Also, in the sessions with patients, it was possible to see that improvements in the interface were necessary in order to improve the patient experience. Before carrying out the first session, it was already observed that not all the classifications should be mapped to a sound. The machine learning program is constantly sending messages with the detected class, causing many sounds to be played in a short time, generating hearing discomfort; therefore, a sound is only reproduced when the detected class is different from the previous one. So, it was observed that it was necessary to refine the interface so that it does not double trigger sounds when performing the movement; that is, when the algorithm misclassified a gesture, we heard an initial note, quickly another note, and then we heard the initial again. This was fixed by requiring a minimum time between heard notes (0.5 seconds). It was also modified during the period of sessions in the hospital that the assignment of musical notes to each class could be chosen to improve the desired musical instrument.

4.3.3 Healthy patients evaluation

The user experience of the interface in people who have not suffered any stroke, allows us to see if the interface of a digital musical instrument can be compared to a traditional instrument. Analyzing the results, we can see that the first two graphics (green) indicate that it takes less practice and it is less tiring to use the digital interface compared to a real instrument, so it is easy to learn to use it and it is not very tiring. This second result (tiredness) could be changed if the people answering the questionnaire were post-stroke patients. The following question only shows a trend that must be taken into account when continuing to develop this digital interface, all the users have preferred the rhythmic accompaniment since they consider that it can be better coordinated with the song. The melodic accompaniment is more difficult to have a relationship with the song because the notes mapped with the gesture may not have any relationship with the melody that sounds. Users in general have found a relationship between the gesture they made and the sound they heard; however, it

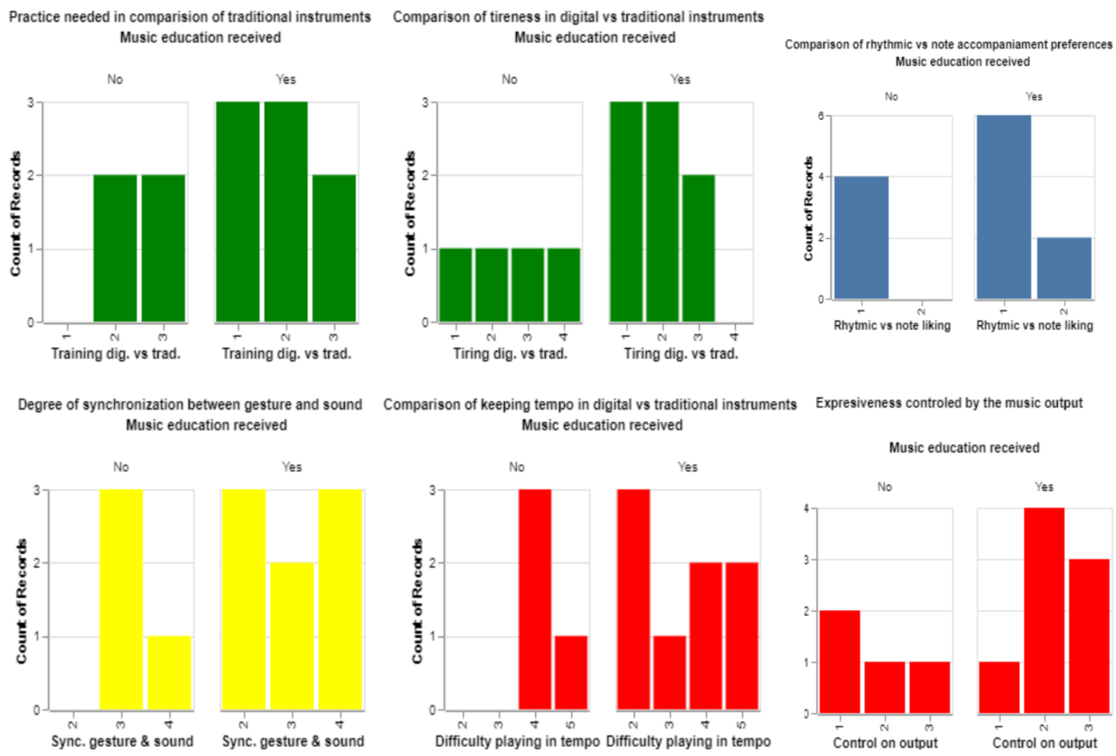

Figure 9: Healthy Patients' Evaluation Graphics

was not very clear and some of them commented that it was difficult to know exactly what gesture to make. Finally, most users have found it quite difficult to keep the tempo of the song and that the interface allowed little musical expressiveness.

Chapter 5

Discussion

5.1 Digital music interface

As we have seen in the methodology, two interfaces have been created, one of them more developed that has been used in the hospital with patients and with users who have tested the interface. The interface that we have finally used has its benefits since it is less tiring than a traditional instrument, it allows a wide range of input gestures and there is a strong relationship between the gestures and the output sounds. The interface seems to have more captivation when the accompaniment is rhythmic.

However, there are still many features to improve the interface. It is difficulty to maintain the tempo; the Wekinator algorithm always return a value between the classes; it generates that when you move between two different gestures, before arriving at the exact position of the second gesture, the sound is reproduced. With the same idea of classifying every image that the webcam captures, sometimes there are errors that patients hear. With the Wekinator program, there is a limitation, as it is not possible to have a system that takes into account the confidence of the classification of the classes. It could be possible that if we combine only the programming languages of Python and Processing, that is to say, leave the Processing interface for: sound reproduction, a better interface design and good interaction, and have another interface for training with Python. By using Python, it might be easier to control the training of the classes and ensure confidence in the predicted class. However, this proposal meant generating another interface proposal that, due to the length of the work, it has not been decided to explore.

Finally, the interface does not allow expressiveness. Maybe by adding some way to control dynamic, the user experience would have been better.

5.2 Limitations

In this section, there are exposed some unavoidable limitations that I encountered during the thesis.

- **Hardware resources.** To increase the number of rehabilitation sessions, some time was dedicated to installing the different programs on other computers, so that these computers could always remain in the hospital. For a few weeks, I tried to have the entire program on a 2015 MacBook, but it was finally finally impossible to install some required Python libraries. Later, the entire interface was working on another more modern computer, 2017 MacBook, on which the screen did not work; however, the hospital did not have a second screen available to view the interface. This meant about 3 weeks dedicated to the installation of the interface to another computer, which was finally not achieved, assuming a lack of time for the sessions at the hospital.

- **Available patients.** During a day, there were few patients, 2 or 3, some of whom were not allowed to do the music therapy rehabilitation sessions. Many times, other patients who could come did not show up for rehabilitation due to fatigue or anxiety, leaving a very limited number of sessions that have been carried out.

- **Absence of a physiotherapist and music therapist in the sessions.** One of the biggest limitations that I have found when implementing the sessions is that there has never been a physiotherapist who has been able to evaluate the patient in a more objective way. During the time that the interface was being developed, the help of the music therapist was useful to guide and give some insights on the process; however, during the implementation of the program in the hospital with patients, the professional music therapist was not able to guide the session. The usual thing was an internship student of a music therapy master's degree being present during the rehabilitation session.

- **Wekinator accuracy.** According to a publication [25] that studies the accuracy of the different classification Wekinator algorithm, the more data used to train the models, the better the accuracy. Therefore, there is a trade-off between accuracy and training time. As the training data is obtained at the beginning of the session, the more data we collect, the more time we will spend creating the classes. As has been mentioned in the previous section, all the Machine Learning progress was like a black box; maybe in others implementations of a digital interface, another programming language should be used to program the models.

5.3 Future work

This section for future work includes different aspects that could continue to be studied and worked on to improve tje interface that has been proposed.

5.3.1 Complete the study

The first point where improvement should be made is to carry out a more complete study, one that is more serious and that therefore the results can be considered reliable and robust. For this, it would be important to implement the experiment that had been designed from the beginning and that it has not been possible to implement in this case.

Longer period of sessions

It is important to increase the session period to have the necessary time to do a minimum of eight sessions on each patient. Taking an extended period of time is important to take into account all those sessions where the patient does not come. In this first implementation, half of the time, the patient did not come to the music therapy and rehabilitation session. By taking a longer period of time, it is possible to really study what the evolution of the patient has been.

More patients control vs. experimental group

Having two groups to do the study, one undergoing rehabilitation sessions and the other undergoing rehabilitation and music therapy sessions, is completely necessary in order to make the study more complete. The comparison of the evolution between these two groups is what will allow us to see the effectiveness of the program and whether the main objective of the interface is met or not. If the evolution of the stroke patients in whom music therapy is included at the time of physiotherapy is more favorable than the evolution of the other group, it means that the use of the application is a success. Since we do not have these two groups, we can not draw any objective conclusions.

5.3.2 Improve the cognitive and musical game

Two of the tasks that the interface have allow to do rehabilitation using a game. In the game, images with musical notes or drums are falling randomly; the game tells you to pick up a note or drum. Using two trained gestures that generate a movement to the left or right you can pick up the images and score points. It would be interesting to improve the cognitive aspect by changing the randomness of the game. If instead of the sequence of notes being random, the sequence follow the notes of a known melody or it is always the same created melody, it will allow the user to expect which will be the next note and thus generate a reward and positive stimuli for the user, which can result in engagement and motivation in using the game and improve rehabilitation.

5.3.3 Increase versatility

It is also important to work on an application that is more portable and versatile. This application currently only works on computers where different elements have been previously installed and where different files are located. These elements are Python and all its libraries, Wekinator, Processing and its libraries, some songs to be able to use if there is no accompaniment from the music therapist and the images used in the interface. In addition, knowledge of how to use the entire system and how to configure Wekinator is needed, something that is not so simple and requires a previous explanation. All this makes the application not so easy to use in other medical centers without the necessary previous installation. So, for the application to be widely used, it is necessary to think of a way to make it much more versatile, portable and easy to use, while also thinking about its implementation on other devices such as mobile phones or tablets.

5.3.4 Avoid delay

The delay that existed between the input gesture and the output sound decreased a lot when using the final (second) proposed interface; however, the delay was not completely removed. The delay can come from different sources, such as the webcam frame rate or the audio system of the computer, so by investigating new computers, maybe the delay decreases more. Another solution, could be exploring a faster language program, such as C.

Chapter 6

Conclusion

This thesis concluded with the creation of a digital music interface for stroke patients' rehabilitation. This interface is close to work in real time, but there is still a slight delay generated maybe by the webcam frame rate, by the audio processing system of the computer or by the language program used, meaning that the system is close to achieving the first objective: it **works in real-time**. The second objective was to have a **personalized interface**. On the interactive interface, by setting the appropriate parameters to detect the body points for each patient assisted by a music therapist or other professional available, the interface can be used for each person in an exclusive way. This interface is also using the KNN machine learning algorithm (with the option to change the classification algorithm in Wekinator settings) to **learn in real-time the user's capabilities**. Sometimes the algorithm predicts the wrong class and generates some errors; the problem is that they are very noticeable, but at least most of the time the classification is correct. Finally, the system has to improve its **evaluation and validation methods**. While the machine learning algorithms had been evaluated and analyzed using precision metric on Python scripts, the interface uses Wekinator machine learning algorithms. The use of two independently algorithms might lead to inconsistencies in the results. In this research, it is assumed that if the best algorithm found for accuracy in Python accuracy is KNN, the classification algorithm used in Wekinator should also be KNN. Another important part of the project is measuring the progress of patients. It has been impossible to use the *Fugl-Meyer* scale or another numeric measurement on the rehabilitation session to have a quantitative evaluation of the evolution; however, some qualitative subjective results have been presented to analyze the patients and the interface. The qualitative results were the comments that the music therapy student wrote after each one of the sessions. The results for one patient showed some improvements in the coordination of movements and in breathing control when performing the exercises.

On the introduction part, there were also two research questions. The first research question is: *Is the combination of music and artificial intelligence useful for stroke rehabilitation?*. The answer is affirmative. According to studies [9] by applying music therapy to sessions, the rehabilitation task becomes more engaging, creating some autonomy for the patients as they learn how to structure the movements. AI benefits this autonomy for rehabilitation because it learns how to adapt the interface with each patient. However, on this interface, it is necessary to study the question for a longer period of time to ensure its effectiveness on the progress of patients. The algorithms used should try to prevent misclassifications and see the real effectiveness in improving the health of stroke patients. The second question addressed is: *Is the artificial intelligence digital music interface comparable to a musical instrument while doing stroke rehabilitation?*. The answer to this question is unclear. On the one hand, the digital music interface seems to be less tiring and easier to play than a traditional music instrument and the percussion sound seems to help more users. However, there are some flaws that distance it from being close to a traditional instrument; these are the lack of expressiveness of the interface and the difficulty of keeping the tempo of a piece due to delay. This means that improvements to the musical content need to be done.

List of Figures

1 Digital music interface system training pipeline. 19

2 Digital music interface system testing pipeline. 19

3 Rehabilitation system workflow. 20

4 First proposal Python interface 27

5 Second proposal Processing interface 28

6 Rehabilitation task visualization 29

7 Rehabilitation game task visualization 30

8 Music Therapists' Evaluation Graphics 35

9 Healthy Patients' Evaluation Graphics 38

List of Tables

1 Paramater settings on Wekinator 17
2 Comparison of interface proposals 31
3 Accuracy of each algorithm using body gestures using different number of classes 33
4 Accuracy of each algorithm using hand gestures using different number of classes 33
5 Body validation & test accuracy of each algorithm 34
6 Hands validation & test accuracy of each algorithm 34
7 Summary of patients' diagnosis and evolution of sessions 36

Bibliography

[1] Association, A. M. T. Music therapy (2023). URL https://www.musictherapy.org.

[2] WHO. Stroke, cerbrovascular accident (2023). URL https://www.emro.who.int/health-topics/stroke-cerebrovascular-accident/index.html.

[3] SEEIUC. The benefits of music therapy in critical patients (2020). URL www.elsevier.es/ei.

[4] Lung, N. H. & NIH, B. I. What is a stroke? (2023). URL https://www.nhlbi.nih.gov/health/stroke.

[5] Antonia Renú, X. C. F., Arturo Urra. Stroke | portalclÍnic. URL https://www.clinicbarcelona.org/en/assistance/diseases/stroke.

[6] Music therapy (2023). URL https://barcelona.guttmann.com/en/treatment/music-therapy.

[7] Denslow, E. Music therapy for stroke patients: How it benefits recovery (2022).

[8] Ghai, S. *et al.* Neurophysiological changes induced by music-supported therapy for recovering upper extremity function after stroke: A case series (2021).

[9] C. Xu, Z. S. F. H., Z. He. Potential benefits of music therapy on stroke rehabilitation (2022).

[10] Grasielle, A. *et al.* Genvirtual: An augmented reality musical game for cognitive and motor rehabilitation.

[11] Baratè, A., Korsten, H. & Ludovico, L. A. A music tangible user interface for the cognitive and motor rehabilitation of elderly people. vol. 2022-October, 121–128 (Science and Technology Publications, Lda, 2022).

[12] Kirk, P. *et al.* Motivating stroke rehabilitation through music: A feasibility study using digital musical instruments in the home. 1781–1785 (Association for Computing Machinery, 2016).

[13] Ambros-Antemate, J. F. *et al.* Accuracy of machine learning algorithms for the assessment of upper-limb motor impairments in patients with post-stroke hemiparesis: A systematic review and meta-analysis. vol. 31 (Wroclaw University of Medicine, 2022).

[14] Ballester, B. R. *et al.* Counteracting learned non-use in chronic stroke patients with reinforcement-induced movement therapy. *Journal of NeuroEngineering and Rehabilitation* **13** (2016).

[15] Weber, R. Z., Mulders, G., Kaiser, J., Tackenberg, C. & Rust, R. Deep learning-based behavioral profiling of rodent stroke recovery. *BMC Biology* **20** (2022).

[16] Opacua, M. I. D. Master thesis on cognitive systems and interactive media emg / motion capture-based accessible music interfaces for rehabilitation (2020).

[17] Mathis, A., Schneider, S., Lauer, J. & Mathis, M. W. A primer on motion capture with deep learning: Principles, pitfalls, and perspectives (2020).

[18] Bishop. Pattern recognition and machine learning.

[19] Scikit-Learn. Support vector machines (2023).

[20] Scikit-Learn. Decision tree (2023).

[21] Scikit-Learn. Ada boost (2023).

[22] Scikit-Learn. Naive bayes (2023).

[23] O'modhrain, S. A framework for the evaluation of digital musical instruments. *Source: Computer Music Journal* **35**, 28–42 (2011). URL https://about.jstor.org/terms.

[24] Vamvakousis, Z. & Ramirez, R. The eyeharp: A gaze-controlled digital musical instrument. *Frontiers in Psychology* **7** (2016).

[25] Rhodes, A. R. . C. R., C. New interfaces and approaches to machine learning when classifying gestures within music (2020).

Appendix A

GitHub Repository

The code used for the digital music interface is available on a GitHub repository. The content of the repository is explained in detail on the README file on GitHub.

The following link allows you to open the GitHub repository: Rehabilitation Interface Repository. The link contains a list of files and folders used for the digital music interface. You can explore the code files, images, drum sounds, the presentation thesis and the paper thesis.

To use the repository, you should clone it to your local computer. Once that all is cloned, it is necessary to generate the executable files using the console (convert from .spec to .exe) as it is explained on section 3 of the README.

For more information, feel free to Raquel Lucena at raquel.lucena01@estudiant.upf.edu.