Motion Initiated Music Ensemble with Sensors for Motor Rehabilitation

Dania Murad[1] **Fenyi Ye***[2] **Michael Barone**[1] **Ye Wang**[1]
[1]*School of Computing, National University of Singapore, Singapore*
[2]*Fudan University, Shanghai, China*
daniamurad@comp.nus.edu.sg, wangye@comp.nus.edu.sg

***Abstract*—Advances in inertial sensors technology used in smartwatches make it possible for movement-based therapies to be both therapeutic and evaluative at the same time. Therapeutic Instrumental Music Performance (TIMP) uses instrument playing for motor rehabilitation. However, applications of TIMP are not generalizable to a variety of motor disabilities and do not often consider the benefits of social interaction. To address the missing opportunities and potential strengths of an integrated technological and therapeutic application, we introduce *MIMES* (Motion Initiated Music Ensemble with Sensors). *MIMES* uses commercially available smartwatches and hand gesture recognition techniques to trigger instrumental sounds with simple hand gestures and support multiple users. We developed this system to provide rehabilitation programs and music therapy for a variety of individuals with motor impairments to improve both physical and mental health outcomes. We discuss *MIMES* system architecture, gesture recognition algorithm, future experimental design and a self-evaluation module to test the efficacy of the technology for rehabilitation.**

***Keywords*-Rehabilitation; Motor Impairment; Music Therapy; Inertial Sensors Technology; Gesture Recognition**

I. Introduction

Harnessing wearable technologies as a tool for rehabilitation offers unique advantages over traditional services including the ability to provide treatment opportunities with remote monitoring for physicians [1]. In particular, technological advances in inertial sensors have progressed to the point where automated gesture recognition systems can be used for motor rehabilitation i.e. restore functions due to disease or injury to the nervous system. For example, Cheol Jeong and Finkelstein [2] used watch-like devices to monitor hand and arm movement with a user-friendly graphical dashboard to motivate exercise and performance. Markopoulos et al. [3] also designed a watch-like device for stroke patients and provided visual feedback to users to improve motivation, self-efficacy, and compliance. However, these applications provided only visual feedback to the user; therapies that incorporate additional sensory modalities such as auditory feedback could further improve health outcomes for individuals with motor dysfunctions [4]. Specifically, motor rehabilitation using inertial sensors with audio-visual feedback would be most appropriately applied within Neurological Music Therapy (NMT).

NMT uses music as an aid to motivate and restore sensory dysfunctions [5] by stimulating the musical perception and production areas of the brain. Furthermore, music therapies have the additional benefit of improving mental health outcomes such as anxiety and depression, which often develop as a result of physical trauma [6]. Some research has already been conducted to explore whether wearable technology can be used for NMT. Newbold et al. [7] developed a wearable device which incorporates music as part of the rehabilitation and evaluates the range of motion for stretching exercises, thus providing a proof of concept for using concurrent audio feedback with wearable devices for music therapy. The system presented in [7] give users an auditory feedback by producing pleasing sounds to facilitate the stretching activity for the patient and displeasing sounds if the user is overexerting during the activity. In the field of NMT, there is a subfield known as Therapeutic Instrumental Music Performance (TIMP) [8], which uses the playing of musical instrument for rehabilitation of motor skills. Given that playing an instrument requires fine motor control, a TIMP session can address many important aspects of rehabilitation including range of motion, muscle coordination and stiffness reduction, hand grasp strength, wrist rotation, stretching exercises and movement of the upper and lower limbs. Musical experience with rhythmic auditory cuing has been demonstrated to reduce the perception of pain, fatigue, and exertion during rehabilitation [4].

Perhaps the related research to digital instrument playing for rehabilitation was conducted by Kirk et al. [9]; digital instruments were used to motivate stroke patients to improve the effective use of their arms. Three stroke patients participated in 15 researcher-led music-making sessions and showed significant improvements in self-management and motor function measures. However, participants performed on digital drum kits which require functional use of the hands and wrists to hold drumsticks, limiting the generalizability of this application for many individuals. Additionally, all of the assessments required clinical evaluations by a health care professional, making the use of the system less feasible for at-home use and self-evaluation. A system which can generate instrumental sounds without the requirement of fine motor control of any type, provide self-evaluation metrics, and promote group collaborative therapy through extrinsic feedback could be beneficial to the health care community by increasing the possible target population. This has the added benefit of easing a health care professional's time resources and improving mental health outcomes for the users.

To address these challenges, we introduce Motion Initi-

*This work was done while Fenyi Ye was an exchange student at National University of Singapore.

ated Music Ensemble with Sensors (*MIMES*), a computer-based music generating system that uses smartwatches for solo or collaborative music composition. The motivation for developing the system is to promote exercise for populations with motor impairment and to improve motor function for instrument playing into simplified hand gestures. Furthermore, with *MIMES* users can practice at home and receive musical feedback as a reward. In the following sections, we discuss the neural network architecture for gesture detection, collaborative performance system, clinical testing paradigm, and future work developing self-evaluation dashboards for users.

II. System Design

There are many classification architectures that are optimal for specific tasks. For instance, hidden markov models (HMM) and support vector machines (SVM) have been used extensively for gesture recognition. For classification and prediction of hand gestures, we have used artificial neural networks (ANN) [10]. For gesture recognition, neural networks take an input signal (raw accelerometer values across 3 axis) and attempt to classify the information into one of a set of pre-defined gestures (outputs). A major benefit of neural networks is that after training the system to recognize a specific gesture, these gestures can then be mapped to any sound, thus producing audio feedback to the user after a gesture is made in real-time. For the current version of *MIMES*, we define nine gestures: flick up, flick down, flick left, flick right, forward, backward, clockwise, counter-clockwise, and swipe down. Figure 1 describes the *MIMES* processing workflow. Details regarding each module are provided below:

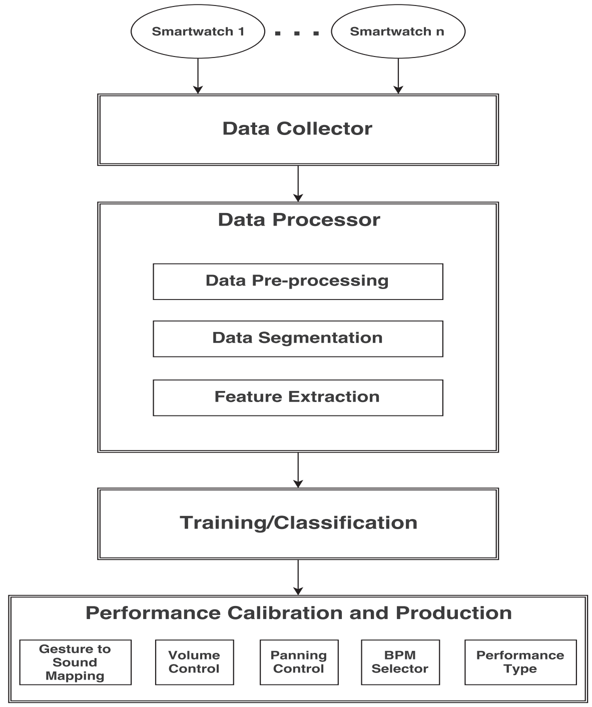

Figure 1. MIMES Architecture Workflow. Each box in the pipeline represents a module: i) *Data Collector* collects accelerometer data from smartwatches via Bluetooth 4.0; ii) *Data Processor* conducts signal pre-processing and feature extraction; iii) *Training and Classification* learns and recognizes gestures using a feed-forward neural network, and; iv) *Performance Calibration and Production* maps gestures to sounds, and determines performance type (synchronized or non-synchronized).

A. Data Collector

The inertial sensors in the smartwatches transmit three-dimensional accelerometer signals (a_x, a_y, a_z) to a computer at a rate of 100Hz via Bluetooth 4.0 and can support up to five devices per system. The module collects these signals from each sensor and transmits them to the Data Processor module for signal filtering and extraction.

B. Data Processor

Three-dimensional accelerometer signals pass through a series sub-processes for real-time detection of gestures. First, high-frequency noise is removed from the accelerometer data using a moving average low-pass filter. This filter eliminates disturbances caused by unintended movements and is expressed as:

$$a[n] = \frac{1}{2M+1}\sum_{m=-M}^{M} a[n+m], \tag{1}$$

where $a[n] = (a_x[n], a_y[n], a_z[n])$ is the acceleration at the n^{th} sample of the data sequence $AS = a[1], a[2], ..., a[L]$ and M is the moving average window size. We use a window size of five sample points ($M = 2$) in the filter per iteration.

In order to provide a natural and seamless experience, the users must be able to perform gestures at any point in time during the performance. This requires a continuous monitoring of meaningful gestures from the filtered signal. The best way to achieve this goal is to detect the sudden change in accelerometer sensor values [10]. For this purpose, a segmentation technique using Euclidean distance is utilized:

$$d(a[n], a[n-1]) = \sqrt{\sum_{i=1}^{n}(a[n]_i - a[n-1]_i)^2}, \tag{2}$$

where $a[n]$ is the accelerometer values across the three axis at the n^{th} sample. Euclidean distance $d(a[n], a[n-1])$ is measured between $a[n]$ and $a[n-1]$. A high value of $d(a[n], a[n-1])$ indicates that a meaningful gesture is performed, thus providing segmented motion interval.

After segmentation, central tendency statistics and time domain features are extracted as gesture features. Feature extraction derives motion characteristics necessary for accurate and meaningful gesture classification. Each feature is calculated across the three axes, generating at least 3 values for the input vector:

Mean: Describes the speed with which the user performed the gesture [11]. It provides 3 features, each for one axis.

Standard deviation(S.D.) and variance: It reveals how much speed is controlled by the user while performing the gesture. It provides 3 features for S.D. and variance respectively.

Minimum and maximum value: Provides the minimum and maximum values in the set of accelerometer values. It helps to distinguish between gestures based on orientation. It also provides 3 features for minimum and maximum value respectively.

Dynamic Time Warping (DTW): Cost function is calculated between a pre-recorded reference signal for each gesture to the newly performed gesture across all the three axes. The two signals will have a minimum value for the cost function if they are similar. There are 9 pre-recorded reference signals for each gesture, each signal having 3 axes. When a new gesture is performed, DTW is calculated between each axis of the newly performed signal and each axis of the 9 pre-recorded gestures. These comparisons provide $3N$ features, where N is the number of gestures defined by the user. DTW helps in differentiating between gestures which have the same mean, standard deviation, minimum and maximum values, but different orientations. Since nine gestures have been defined for this iteration of *MIMES*, 27 values are included in the input vector for DTW. A description of DTW can be found in [12].

C. Training and Classification

In the training phase, raw values from gestures are collected and passed through the pre-processing and segmentation technique. 42 salient features, as described in Section II-B, are extracted from the segmented gesture and provided as inputs to the feed-forward neural network consisting of one input, one hidden and one output layer. The input layer contains 42 input nodes, representing the 42 extracted features for nine gestures. The hidden layer consists of 40 nodes and utilizes a sigmoid transfer function. A softmax transfer function is used for output layer which has nodes that represents each of the nine predefined gestures. Lastly, back-propagation is used to train the model.

For our demonstration, the model is trained with a dataset of 1080 gesture recordings (120 recordings per gesture) collected from a single user, with nine gestures (described in Table I) as labels. To test the classification accuracy of the system with one person, each of the nine gestures is executed one hundred times, testing a new gesture each time. A trial is considered successful if the output corresponds to the appropriate gesture with $> 90\%$ confidence, an empirically determined threshold. Table I illustrates the percentage accuracy for each output class. The average accuracy across all classes is 95.9% ($St.Dev = 3.7\%$), achieving a minimum of 89%.

Gesture	Success ($> 90\%$ Confidence)
Flick Up	94%
Flick Down	95%
Flick Right	91%
Flick Left	98%
Forward	89%
Backward	100%
Swipe Down	99%
Clockwise Circle Shaped	98%
Counter-clockwise Circle Shaped	99%

Table I
PRELIMINARY ANALYSIS OF MIMES GESTURE RECOGNITION ACCURACY.

D. Performance Calibration and Production

Once the network has been trained, users can calibrate their composition by mapping any sound to the output classes. Users can also select one of the two modes: non-synchronized and synchronized. In the non-synchronized mode, audio playback occurs as soon as a gesture is performed; this mode is useful for users who want to understand the fundamentals of music playing in a solo performance setting, or for a group of users with a strong sense of rhythm. The non-synchronized mode is also useful for free and improvised performances or for users who have become well-versed in using the system. In synchronized mode, the system automatically corrects the minor errors in timing by adjusting the sound to a certain tempo. It enable users to create a musical piece that adheres to a specific tempo regardless of musical expertise or entrainment impairments. A tactile metronome is provided to the user by making the smartwatch vibrate with a pre-determined beat. This cues users when they should perform a gesture, minimizing asynchrony between them; synchronized mode is effective for learning pre-determined choreography, or for individuals with motor and rhythm impairments. For a natural sounding group performance, a wide range of musical instruments can come together with a keen sense of rhythm, harmony, and melody. For this purpose, the amplitude and spatial arrangement of each element are also incorporated into the framework. The music produced by the performed gestures then renders the mixed musical audio signal stream, a combination of different instruments from different smartwatches, to the output device (speakers) in real-time.

III. FUTURE WORK

We will test the efficacy of *MIMES* as a motor rehabilitation tool by preparing a TIMP intervention class and letting elderly individuals with motor impairments use sensors. We base our experimental design off of the work of [13], who conducted music therapy intervention trials on people with Parkinson's disease. The treatment will involve 32 participants split into two groups, one group receiving the *MIMES* intervention (test), the other receiving a traditional physical therapy treatment (control). Both groups will be balanced concerning motor impairment, age, sex, quality of life, physical health, and mental well-being. Interventions will be 1.5 hours, with four tasks over the course of 14 weeks: i) 15 minutes for participants to have open, solo play to re-familiarize themselves with the system; ii) 30 minutes of group rehearsal where participants will perform a routine developed by the researchers as a group; iii) 15 minutes for unstructured group performance, where users can perform together without restriction and; iv) 30 minutes for bi-weekly rotational assessment by therapists or the self-evaluation module described in Section III. Tasks will be randomized to control for ordering effects. To assess whether social tasks are improving mental health outcomes, participants will be required to fill out surveys regarding quality of life (Health Related Quality of Life Questionnaire and Caregiver's questionnaire) and depression (Beck's Depression Inventory).

Self-Evaluation Using Smartwatches A significant time investment is required for evaluating motor dysfunctions. Health care practitioners must assess motor functioning for each patient manually with standardized clinical assays. The movements defined in current system can only be used to motivate people to stay fit. However, it can be extended to incorporate clinically approved movements for rehabilitation. Self-evaluation using smartwatch technologies have already been considered for Parkinson's Disease detection [14], therefore an evaluation module that analyzes movements used for clinical assessment of other types of movement impairments would be valuable for both patients and practitioners. To maintain the motivating aspects of movement with music, the module is being developed with a user-centric approach with elements of gamification. The user-centric approach simplifies how users interact with the module, helping elderly populations to use the technology without confusion.

The self-evaluation module will instruct users to perform a series of movements from a seated position which occur in daily life. Movements will be similar to those evaluated in our experimental design which includes daily life activities. Inertial sensors will process movement data similar to the *Data Collection and Data Processor* processes in the performance application. Results will be compared to past evaluations to determine if the user is improving over time. The user can be informed of their performance as well as where they need to improve. Since the health care practitioner can monitor performance remotely, it will be easier for them to determine if the user's movement impairments gets worse or plateau.

IV. Conclusion

Irrespective of the outcomes found from our experimental testing, our aim is to demonstrate that building *MIMES* highlights the benefits of incorporating commercially available technologies. Our endeavor aspires to provide tools that give autonomy to the patient for self-rehabilitation, while reducing the time investment of caregivers. We hope that *MIMES* inspires future work that considers the importance of interventions which improve mental health outcomes and physical trauma through social, holistic interventions that are motivating and joyful.

Acknowledgment

We thank Arun Shenoy for his guidance during the development of the prototype. This project was partially funded by a research grant R-252-000-602-112 from Ministry of Education in Singapore.

References

[1] P. Bonato, "Advances in wearable technology and applications in physical medicine and rehabilitation," *Journal of neuroengineering and rehabilitation*, vol. 2, no. 1, p. 2, 2005.

[2] I. cheol Jeong and J. Finkelstein, "Computer-assisted upper extremity training using interactive biking exercise (ibike) platform," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*. IEEE, 2012, pp. 6095–6099.

[3] P. Markopoulos, A. A. Timmermans, L. Beursgens, R. Van Donselaar, and H. A. Seelen, "Us' em: The user-centered design of a device for motivating stroke patients to use their impaired arm-hand in daily life activities," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*. IEEE, 2011, pp. 5182–5187.

[4] H. A. Lim, K. Miller, and C. Fabian, "The effects of therapeutic instrumental music performance on endurance level, self-perceived fatigue level, and self-perceived exertion of inpatients in physical rehabilitation," *Journal of music therapy*, vol. 48, no. 2, pp. 124–148, 2011.

[5] C. P. Hurt-Thaut and S. B. Johnson, "Neurologic music therapy," *Music Therapy Handbook*, p. 220, 2015.

[6] S. Guetin, F. Portet, M. Picot, C. Pommié, M. Messaoudi, L. Djabelkir, A. Olsen, M. Cano, E. Lecourt, and J. Touchon, "Effect of music therapy on anxiety and depression in patients with alzheimers type dementia: randomised, controlled study," *Dementia and geriatric cognitive disorders*, vol. 28, no. 1, pp. 36–46, 2009.

[7] J. W. Newbold, N. Bianchi-Berthouze, N. E. Gold, A. Tajadura-Jiménez, and A. C. Williams, "Musically informed sonification for chronic pain rehabilitation: Facilitating progress & avoiding over-doing," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2016, pp. 5698–5703.

[8] M. H. Thaut and V. Hoemberg, *Handbook of neurologic music therapy*. Oxford University Press (UK), 2014.

[9] P. Kirk, M. Grierson, R. Bodak, N. Ward, F. Brander, K. Kelly, N. Newman, and L. Stewart, "Motivating stroke rehabilitation through music: A feasibility study using digital musical instruments in the home," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2016, pp. 1781–1785.

[10] R. Xie and J. Cao, "Accelerometer-based hand gesture recognition by neural network and similarity matching," *IEEE Sensors Journal*, vol. 16, no. 11, pp. 4537–4545, 2016.

[11] F. Hong, S. You, M. Wei, Y. Zhang, and Z. Guo, "Mgra: Motion gesture recognition via accelerometer," *Sensors*, vol. 16, no. 4, p. 530, 2016.

[12] E. Keogh and C. A. Ratanamahatana, "Exact indexing of dynamic time warping," *Knowl. Inf. Syst.*, vol. 7, no. 3, pp. 358–386, Mar. 2005. [Online]. Available: http://dx.doi.org/10.1007/s10115-004-0154-9

[13] C. Pacchetti, F. Mancini, R. Aglieri, C. Fundarò, E. Martignoni, and G. Nappi, "Active music therapy in parkinsons disease: an integrative method for motor and emotional rehabilitation," *Psychosomatic medicine*, vol. 62, no. 3, pp. 386–393, 2000.

[14] V. Sharma, K. Mankodiya, F. De La Torre, A. Zhang, N. Ryan, T. G. Ton, R. Gandhi, and S. Jain, "Spark: personalized parkinson disease interventions through synergy between a smartphone and a smartwatch," in *International Conference of Design, User Experience, and Usability*. Springer, 2014, pp. 103–114.