A PRELIMINARY EVALUATION OF A SMART EXERCISE APPLICATION WITH WEARABLE SENSORS

Eric Vasey[1], *Xingliang Li*[1], *Tejin Yoon*[2], *Myounghoon Jeon*[1]

Michigan Technological University
Mind Music Machine Lab[1], Neuromechanics Laboratory[2]
1400 Townsend Drive, Houghton, Michigan, United States
{emvasey,xli17,tyoon,mjeon}@mtu.edu

ABSTRACT

When doing lower leg exercises for strength training or to properly recover from a leg injury, people can encounter problems with boredom, lack of feedback on exercises, and potentially a worse patient outcome or incomplete recovery. We have previously developed and validated a Smart Exercise application with a wearable Bluetooth IMU, and we are now investigating the usability of the system and the usefulness, intuitiveness, and appropriateness of the auditory and visual feedback mechanisms. Our results indicate that the system has effective, intuitive, and appropriate auditory and visual feedback and shows potential for future improvement.

1. INTRODUCTION

The repetitive nature and lack of feedback while doing lower leg exercises for strength training can cause boredom and lower motivation to exercise. In the case of an individual with an injury, completing at-home physical therapy exercises is important for patients to make a full recovery. However, the boring, repetitive nature of these exercises and lack of feedback on performance can lower patient compliance and result in worse patient outcomes. In addition, physical therapists do not have access to any data on patient compliance or know if patients are performing their exercises correctly.

To address these issues, we have developed a Smart Exercise application which combines an Android phone and a low-cost Bluetooth IMU (Inertial Measurement Unit) from *Mbientlab* to track performance in forward lunges and squats while delivering real-time audiovisual feedback to improve a user's performance and motivation to exercise. Data are collected from the IMU and streamed to the phone at 100 Hz from the accelerometer and gyroscope, and at 30 Hz from the magnetometer. The data are passed through a threshold based filter that allows data through that exceeds 5% of at-rest calibrated values, then a Kalman filter is applied to remove sensor noise from the data.

Noh et al [1] conducted a pilot study of 13 participants to evaluate the functionality of the application and validate the data with a research grade accelerometer. The application performed as anticipated for detecting forward lunges and measuring participant balance on a one-leg half squat task. Pearson's correlation coefficients were computed for the accelerometer and gyroscope data, and indicated a high correlation between our IMU and the research grade IMU (r = 0.90 to 0.99). A Bland-Altman plot was also utilized to validate the data, and indicated a low discrepancy between the IMUs.

1.1. Related Works

Spina et al [2] developed and evaluated an Android phone application to assist the fitness training and rehabilitation of patients with COPD (Chronic Obstructive Pulmonary Disease). The application utilizes the Android phone's on board accelerometer and magnetometer to generate an exercise model that personalizes the system to a patient and allows the patient to later complete exercises that are evaluated with regard to the prior model. The app provided auditory feedback for incorrect repetitions of an exercise, along with suggestions to improve. Evaluation indicated that exercise repetitions were classified at 96.7% accuracy, and the auditory feedback was effective in guiding participants to perform exercises correctly. However, this application required the phone to be on the user's arm, limiting the usefulness of visual feedback.

Rosati et al [3] investigated the impact of task-related, error-related, and visual-related feedback on a motion tracking exercise using a two degree of freedom joystick and touchpad. Their findings indicated that auditory feedback was more beneficial for the upper-body exercise task, visual task-related feedback improved performance significantly more than visual error-related feedback, and that auditory feedback on the motion of the target was better than feedback from moving the interactive device.

2. PROCEDURE

In this study, we conducted an evaluation of the usability of the Smart Exercise application and the appropriateness, usefulness, and intuitiveness of the application's auditory and visual feedback in its Lunge Piano/Lunge Sequence function. This function had users perform lunges at seven equally sized regions on a half-circle, with each region corresponding to a note on a C scale. Auditory feedback was given in the form of the corresponding note being played on the phone's speakers, and the region the user was facing was highlighted on a half-circle visualization (Figure 1). Lunge Sequences were series of notes that indicated the regions that the user needed to perform a lunge at in a particular order. These sequences were randomly generated and communicated to the user via audio and visual feedback. For example, if the sequence was C-D-G-E, the user would need to face the region for C, perform a forward lunge, then rotate to face D, and perform a lunge, proceeding in this manner until they have performed lunges while facing all of the indicated directions in the correct order.

Audio feedback consisted of a text-to-speech voice instructing them which sequence notes/regions to perform lunges at. Performing a lunge at a correct region would play the corresponding note. Successfully completing a sequence was immediately followed by a preview of a new sequence. Performing a lunge at an incorrect region would play the corresponding note, then a series of three 100ms 311Hz tones that are 100ms apart, followed by a text-to-speech voice indicating that they were incorrect. Visual feedback was a half-circle visualization in the app which illuminated sections of the view corresponding to the notes/directions to perform lunges at. Performing a lunge at the correct region caused that region to fill with red, yellow, magenta, gray, cyan, light gray, or green depending on the region. Successfully completing a sequence would display 'Great Job! You completed the sequence correctly". Performing a lunge at an incorrect region would display "Sorry! That's not quite correct".

Initially, participants were asked to complete the self-administered short version of the International Physical Activity Questionnaire (IPAQ). Then, they were asked to complete three sequences of four notes under three different feedback conditions - audio feedback only, visual feedback only, and both audio and visual feedback with order counterbalanced. In the audio feedback condition, the phone was taken from the participant, and they had to rely on their memory of the sequence and the ground tape markers (Figure 2) to indicate their direction. In the visual condition, the phone volume was muted, forcing the participants to utilize the visualization and messages to communicate the sequence to them and indicate they had completed a lunge correctly. If a participant performed a lunge at an incorrect note in the sequence, all subsequent notes were marked as incorrect, since the lunges needed to be performed in the correct order. Then, the error would be communicated to the participant, and a new sequence would be generated and provided to the participant. The app would log the sequence as well as the number of correct notes. After completing the three sequences under the three conditions, participants were asked to complete a questionnaire containing System Usability Scale (SUS) and the experimenter-designed questions.

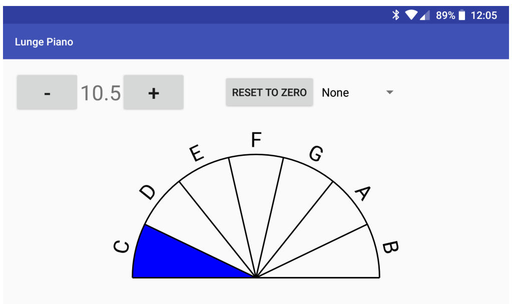

Figure 1: Lunge Piano/Lunge Sequences Visualization.

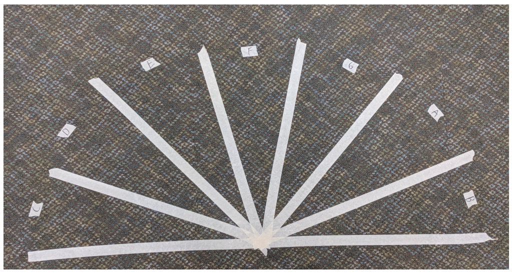

Figure 2: Lunge Piano/Lunge Sequences Floor Markers.

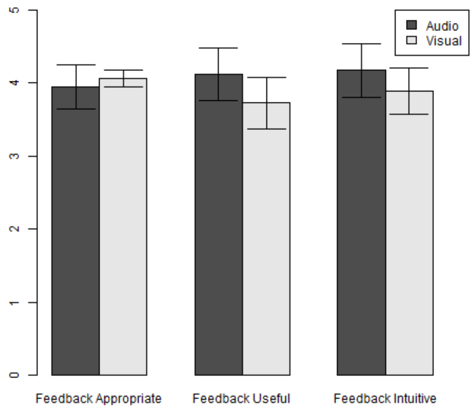

Figure 3: Mean appropriateness, usefulness, and intuitiveness of audio and visual feedback.

2.1. Questionnaires

2.1.1. System Usability Scale (SUS)

The System Usability Scale consists of ten questions scored on a five-point Likert scale that aim to evaluate the usability of a system [4]. This is intended to evaluate the usability of the system as a whole.

2.1.2. International Physical Activity Questionnaire (IPAQ)

IPAQ is a questionnaire that is used to evaluate the physical activity level of an individual over a given time period and categorize it into low, moderate, or high levels [5]. We utilized the self-administered short version of IPAQ, which surveys physical activity over a seven day period.

2.1.3. Experimenter-designed Questionnaire

To investigate the app's feedback modalities, participants were asked to rate the audio and visual feedback mechanisms in terms of their usefulness, intuitiveness, and appropriateness for the given task.

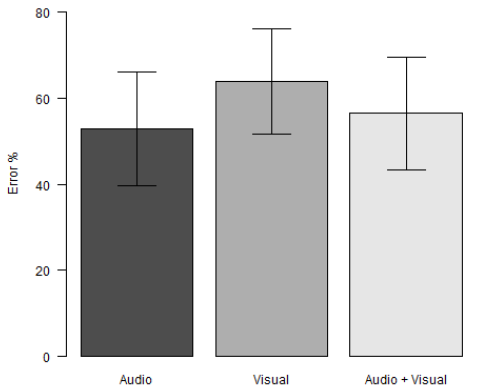

Figure 4: Mean percent error for the Audio, Visual, and both Audio and Visual conditions.

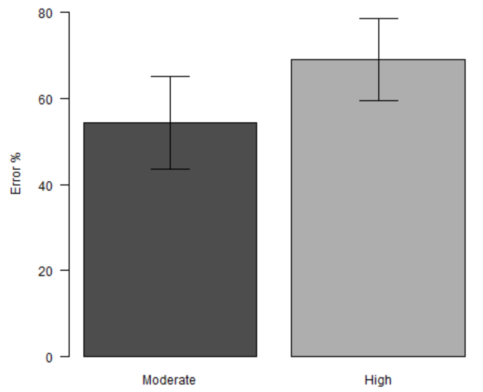

Figure 5: Mean percent error for Moderate and High activity individuals.

3. RESULTS

A total of 18 participants (5 female, M = 19.72, SD = 1.18) completed the above procedure. The ten questions from the SUS were scored using the criteria in [4] (M = 61.39, SD = 15.86), and IPAQ was scored using the criteria at [6]. Five participants were excluded from IPAQ scoring due to incomplete responses. Of the 13 remaining, 6 were classified as having high levels of physical activity, and 7 as having moderate levels of physical activity. A paired samples t-test did not indicate any significant differences between the perceived appropriateness (t = -0.809, $p < 0.43$), usefulness ($t = 2.051, p < 0.056$), or intuitiveness ($t = 1.686, p < 0.11$) of the audio or visual feedback (Figure 3). A two-way repeated measures ANOVA on the mean percent error (Figure 4) did not indicate any significant differences (F(2,30) = 2.355, $p < 0.112$) between the audio only (M = 52.78, SD = 28.01), visual only (M = 63.89, SD = 26.04), or audio with visual conditions (M = 56.48, SD = 27.79). No significant difference in the mean percent error was found (F(1,30) = 0.081, $p < 0.778$) between individuals at a moderate physical activity level vs. individuals at a high activity level (Figure 5). The interaction effect for activity level and condition (Figure 6) was near significance (F(2,30) = 3.187, $p <$ 0.056).

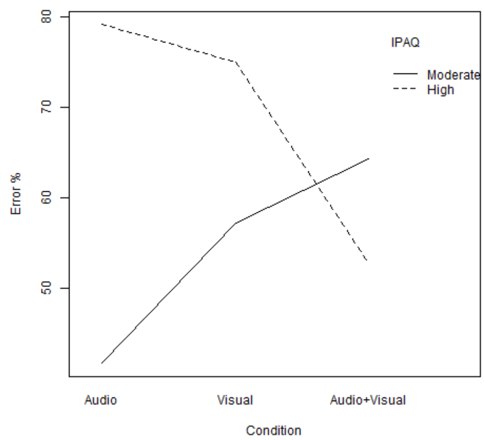

Figure 6: Interaction between IPAQ Physical Activity Level and Condition.

4. DISCUSSION

Utilizing the adjective rating scale developed by Bangor, Kortum, and Miller [7], the application's SUS score is between *OK* and *Good*. This indicates that we are on the right track in terms of the usability of the system, but additional improvements are needed.

While the results are not at the level of statistical significance, they are promising. The mean percent error for the audio only condition was lower overall than visual only, with audio and visual in-between. This may be due to the improved effectiveness of the auditory feedback in indicating a correctly performed lunge. In addition, the policy for errors was harsh - if a participant performed a lunge at any incorrect note or a note in the wrong order, all subsequent notes in the sequence were marked as incorrect, the visual and audio feedback for an incorrect sequence would occur, and a new sequence would be given to the participant. This might also have influenced the SUS scores - participants might have rated the system lower due to its harsh policy on performance.

Participants rated both the visual and audio feedback as appropriate, useful, and intuitive (around 4 out of 5). The differences between auditory and visual feedback in usefulness and intuitiveness are approaching significance. Since the feedback was good overall, this could be due to the ceiling effect. Participants who were moderately physically active performed better with audio only feedback than those who were highly physically active. The more physically active participants may have been more confident in their ability to perform forward lunges correctly using their own technique. This might have impacted their compliance with what the application expected to be an appropriate lunge, and possibly proceeding with the sequence ahead of the application.

5. CONCLUSION

Based on our results here, we have a solid base from which we can improve the overall design and usability of the system. By adding additional tutorials to the application and iterating on the auditory and visual feedback, we can better guide users to perform their exercises more accurately and make them more enjoyable. We intend to conduct a similar evaluation to this one in the future with an updated version of the application and more forgiving failure conditions, and ultimately an in-home extended trial of the system.

6. REFERENCES

[1] B. Noh, E. Vasey, K. Phillips, D. Verbrigghe, M. Jeon, and T. Yoon, "Cost-effective personal training aid to improve leg function using smart exercise application: Pilot study," in *Medicine and Science in Sports and Exercise*, vol. 50, no. 5 Supplement, 2018.

[2] G. Spina, G. Huang, A. Vaes, M. Spruit, and O. Amft, "Copdtrainer: A smartphone-based motion rehabilitation training system with real-time acoustic feedback," in *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '13. New York, NY, USA: ACM, 2013, pp. 597–606. [Online]. Available: http://doi.acm.org/10.1145/2493432.2493454

[3] G. Rosati, F. Oscari, S. Spagnol, F. Avanzini, and S. Masiero, "Effect of task-related continuous auditory feedback during learning of tracking motion exercises," in *Journal of Neuro-Engineering and Rehabilitation*, vol. 9, 2012, p. 79.

[4] J. Brooke, "Sus: A quick and dirty usability scale," https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html, 1996.

[5] C. L. Craig, A. L. Marshall, M. Sjstrm, A. E. Bauman, M. L. Booth, B. E. Ainsworth, M. Pratt, U. Ekelund, A. Yngve, J. F. Sallis, and P. Oja, "International physical activity questionnaire: 12-country reliability and validity," in *Medicine and Science in Sports and Exercise*, vol. 35, no. 8, August 2003, pp. 1381–1395.

[6] T. I. group, "Ipaq scoring protocol," https://sites.google.com/site/theipaq/scoring-protocol, November 2005.

[7] A. Bangor, P. Kortum, and J. Miller, "Determining what individual sus scores mean: Adding an adjective rating scale," *J. Usability Studies*, vol. 4, no. 3, pp. 114–123, May 2009. [Online]. Available: http://dl.acm.org/citation.cfm?id=2835587.2835589