A FOCUS-PLUS-CONTEXT VISUALIZATION FOR STRENGTH EXERCISE TRAINING

Daniel Steffen¹, Florian Kalt¹, Michael Resch² and Gabriele Bleser¹

¹*Technische Universität Kaiserslautern, AG wearHEALTH, Kaiserslautern, Germany*

²*ERGO-FIT GmbH & Co. KG, Pirmasens, Germany*

ABSTRACT

Regular strength exercise training is recommended by the WHO. Users – especially novices or patients – call for intensive supervision to execute strength exercises correctly – even if performed with a fitness device. Usually, a trainer or a physiotherapist provide guidance and assist the user. However, increasing costs and missing experts call for new solutions that support the user to perform strength exercises correctly and without supervision of an expert.

This paper describes the fitness system SMART-FIT. The system comprises a cable machine, a notebook, and a tablet PC providing visual feedback during strength exercise training. 36 participants volunteered in our user study, designed as within-subject experiment. Performing a shoulder exercise, participants tested 3 different visual feedback interfaces: *2-Pillar*, *1-Pillar*, and *Ribbon*. We evaluated the effect of the interfaces on the performance, e.g., Pearson's correlation coefficient. Additionally, we applied the questionnaires Flow-Short-Scale, NASA-TLX, AttrakDiff2, and System Usability Scale among others. The results of our user study will enable designers to make more appropriate design choices for future exercise support systems.

KEYWORDS

Strength Exercise, Focus-Plus-Context, Visual Augmented Feedback, Error Measurements, User Study

1. INTRODUCTION

Physical activity including aerobic and strength training provides many physiological benefits. Regular strength exercise training is recommended by the World Health Organization (WHO) both in prevention and in rehabilitation. It effectively reduces falls and reduces the risk of disease outcomes. However, users – especially beginners or patients during rehabilitation – have to be supervised and supported intensely to ensure correct exercise execution and to avoid infringements. A badly executed exercise can lead to injuries and might delay the recovery. Frequently, a trainer or physiotherapist guides and supports the user in the fitness or rehabilitation center, i.e., a controlled environment. Increasing costs and the absence of qualified personnel call for innovative and flexible solutions that can be used anywhere and even without supervision. Thus, the correct execution of an exercise without supervision is the overall challenge.

There are many fitness devices available in the market, which provide visual feedback to the user through a mounted display (cf. ERGO-FIT 2019). Many of these systems are limited to display only the number of repetitions, the number of sets or the information about breaks. However, additional factors like the range of motion (ROM), the moving amplitude, the speed of eccentric and concentric movements, or the movement smoothness are important variables to guarantee a correct and safe exercise execution.

The problem of executing exercise movements while simultaneously validating the correct execution is a major challenge. Among others, visual feedback can be used to address this problem. Visual feedback for movement tasks can be subdivided into abstract visualizations or natural visualizations (Sigrist et al. 2013). Abstract visualizations – such as lines, curves, bars or points – are being used to provide visual feedback for simple movement exercises. Erikson et al. (2011) evaluated a simple visual feedback for runners. The authors used three bars to visualize the variables vertical displacement, step length, and the mechanical power-consumption, which was a combination of the previous two bars. The height of the bars indicated the values of the parameters at the moment the participant was running. In each bar, a mark was indicating the target level. Morris et al. (2007) evaluated a visualization for learning a sequence of one-dimensional forces.

They used two bars representing target and actual force. Chittaro and Sioni (2014) compared simple abstract visualizations within a mobile breathing app: a circle-based visualization (circle expands and contracts, which suggests the expansion and contraction of human lungs) and a wave-based visualization where a wave-like line illustrates the optimal respiratory cycle.

In contrast, natural visualizations are utilized for more complex movements. A movement is considered complex if the movement consists of several degrees of freedom. Steffen et al. (2011), for example, used wireless movement sensors to track users' movements and to mirror complex exercise movements in VR.

The following sections will describe the SMART-FIT system to support simple movement exercises. We developed 3 training interfaces that belong to the category of abstract visualization for movement exercises.

Figure 1. (a) ERGO-FIT's cable machine uses a distance sensor to measure the displacement of the weights. We added a notebook streaming the distance data via Wi-Fi to a 10 inch tablet PC displaying the training interface. (b) Top view of the internal and external rotation of the shoulder exercise in the user study

2. SMART-FIT

Within this paper we present the SMART-FIT system. The SMART-FIT system is displayed in Figure 1. The main component of the system is a cable machine of the ERGO-FIT GmbH & Co. KG. The cable machine utilizes a high performance distance sensor measuring the displacement of the weight in millimeter. The cable machine is equipped with a display that is mounted on a swiveling arm. We enhanced the system with an extra notebook and a 10-inch tablet PC (same display size as the supplied display). The tablet PC is mounted in front of the supplied display. We connected the distance sensor to the notebook and streamed the data over Wi-Fi to the tablet PC. Firstly, we added the extra hardware components to store the movement data for later analysis. Secondly, we aimed for an architecture that is capable of streaming the exercise data to a physiotherapists' smart phone to monitor multiple users, for example.

Figure 2 presents three different training interfaces to support the exercise execution: *2-Pillar*, *1-Pillar*, and *Ribbon* interface. The maximum and minimum displacements for an exercise, i.e., the maximum and minimum flexion / extension, are computed during an initial calibration routine. They define the users' individual range of motion (ROM). The user should not move above or below the maximum and minimum to avoid any injury. A repetition is completed when the user moves the actual from the lower area to the upper area and back again. The target displacement defines speed and position of the exercise, i.e., the displacement of the weight. All interfaces display the number of repetitions and number of sets to perform / performed. Each interface has a stop / pause button to cancel the exercise at any time.

The *2-Pillar* interface uses two animated pillars (see Figure 2, left). The fill level displays the target and actual weight displacement. The pillars are labeled with 'target' and 'actual' to avoid misunderstandings. Using the *2-Pillar* interface might result in a cognitive overload. On the one hand, the user focuses on a correctly executed repetition with respect to the ROM, i.e., moving above and below a specified threshold. On the other hand, he has to pay attention to align actual and target displacement to perform the exercise at the correct speed. This is a dual-task condition which might result in a weak performance while performing the exercise.

Figure 2. The developed training interfaces: *2-Pillar* (left), *1-Pillar* (center), and *Ribbon* (right). Each interface illustrates maximum, minimum, actual, and target weight displacement. To complete a repetition the user starts below the lower threshold, moves the actual above the upper threshold and back again

To tackle the dual-task condition we developed the *1-Pillar* interface (see Figure 2, center). Target and actual displacement are merged into a single pillar to avoid cognitive overload. The fill level of the pillar illustrates the target displacement. A horizontal bar represents the actual weight displacement.

The idea for the *Ribbon* interface (see Figure 2, right) goes back to wire loop games. Here, the metal loop, i.e., the actual displacement, is a filled circle, which is controlled by the user and moves along the vertical axis. A gray ribbon illustrates the target displacement and moves horizontally from left to right. To perform a correct movement the circle should stay within the ribbon. The *Ribbon* interface is a focus-plus-context visualization. The circle represents the actual displacement, i.e., the temporal focus. The graphical element ribbon visualizes past, current, and future target displacement, i.e., the temporal context. This interface allows the user to perceive the duration of a pause between repetitions. Additionally, the slope of the ribbon provides information about how fast he will have to perform concentric and eccentric movements.

3. USER STUDY

We conducted a within-subject experiment with the independent variable *Training Interface (TI)* with three levels: *2-Pillar*, *1-Pillar*, and *Ribbon*. In order to avoid learning and crossover effects we used a fully counterbalanced setup. To evaluate the performance of the exercise repetitions we used the following error measures as dependent variables: Pearson's Correlation Coefficient (COR), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Russell's Error Measurements (RUS; Russell 1997). Additionally, we used the following questionnaires: Flow-Short-Scale (FSS; Rheinberg, Vollmeyer, and Engeser 2003), Raw NASA TLX (RTLX; Hart 2006), System Usability Scale (SUS; Brooke 1996), and AttrakDiff2 (Hassenzahl, Koller, and Burmester 2008).

In the user study we used the SMART-FIT system as described in the previous chapter. A Samsung Galaxy Tab A10.1 (2016) tablet PC – running Android 7.0 – was used to display the different *TI* conditions. The tablet offers a resolution of 1200 x 1920 pixels.

The participants were students or employees at our University. They were recruited through notices posted on campus or by e-mail. 36 participants (14 female), aged 21 to 66 years ($M = 40.22$, $SD = 15.78$) volunteered in the user study. All but one participant were right-handed. 50 % of the participants have never been to the gym, while 20 % visit the gym on a regular basis. The majority of participants (21) were unfamiliar with exercises that involve a smart fitness device while 15 participants had some familiarity with such smart exercise devices. 2 out of 10 participants own a fitness device. Participants did not receive any compensation. No participant has been excluded from the user study.

The experiment was conducted in our laboratory. At the beginning of the experiment the participants signed the informed consent. Furthermore, we gave some general information and explained how to perform the exercise correctly. Figure 1 illustrates the exercise. It is a typical and simple strength exercise that is recommended in the case of the impingement syndrome of the shoulder. The exercise is composed of an external and internal rotation of the shoulder joint. Depending on the users' dominant hand we configured the hardware set up.

Figure 3. Left: The *Ribbon* interface received the highest mean and median scores in the FSS. Right: Raw NASA TLX (RTLX) scores and standard error bars. No significant differences were found for mental (RTLX-MD), physical (RTLX-PD), and temporal demands (RTLX-TD), performance (RTLX-P), effort (RTLX-E), and frustration (RTLX-F)

Before starting the application, the participants performed the exercise without any interface. By doing this we determined the maximum and minimum displacement of the weight and consequently the participants' individual ROM, i.e., the maximum internal and external rotation. The duration for the internal rotation was set to 3 seconds and the duration for the external rotation was set to 2 seconds. The weight for the exercise was 1kg for all participants. We chose this weight to avoid fatigue and to ensure all participants being able to perform all repetitions.

Before each trial, the participants received a training to familiarize themselves with the selected *TI*. During the training they performed up to five repetitions. Ten seconds after the training had finished the trial phase started automatically. Participants were asked to perform 10 repetitions of the shoulder exercise as correct as possible. The application stopped when 10 repetitions have been performed or the user canceled the application.

After the trials have been finished, we asked participants to rank the *TIs*. Besides their gym habits, participants reported on their demographic variables (e.g. year of birth, gender, etc.). The experimental user study lasted between 31 and 92 minutes ($MD = 45.39$, $SD = 11.42$). The following sections describe the results of the measures taken.

4. RESULTS

4.1 Flow-Short-Scale

Within our experiment we evaluated the effect of the *TI* on the participants' flow state. We used the German version of the FSS questionnaire (Rheinberg, Vollmeyer, and Engeser 2003). The FSS consists of 13 items. Each item is rated on a 7-point Likert scale from 1 ("no agreement") to 7 ("full agreement"). The first 10 items measure the *Flow Experience*. The items 11-13 measure *Worry* someone may have in the situation the measurement is made.

Figure 3 (left) shows the result of the FSS questionnaire. Measuring the participants' flow reveals that the *Ribbon* interface received the highest scores. The average *Flow Experience* using the *Ribbon* interface was larger ($M = 5.62$, $SD = 1.01$) than using the *2-Pillar* interface ($M = 5.35$, $SD = 0.90$) or the *1-Pillar* interface ($M = 5.27$, $SD = 0.92$). The differences between the interfaces conditions were not significant, $F(2, 70) = 2.68$, $p = 0.07$, $\eta^2 = 0.02$.

Furthermore, the factor *Worry* was rated highest when using the *Ribbon* interface ($M = 2.81$, $SD = 1.29$). Participants rated the *2-Pillar* interface lower ($M = 2.33$, $SD = 1.14$) than the *1-Pillar* interface ($M = 2.54$, $SD = 1.34$). Mauchly's test indicated that the assumption of sphericity had been violated, $W = 0.82$, $p < 0.05$, therefore degrees of freedom were corrected using Huynh-Feld estimates of sphericity ($\epsilon = 0.88$). These differences were significant, $F(1.77, 61.86) = 5.92$, $p < 0.05$, $\eta^2 = 0.02$. Bonferroni post hoc tests revealed a significant difference between the *Ribbon* interface and the *2-Pillar* interface ($p < 0.01$). No other comparisons were significant (all $ps > 0.05$).

Figure 4. Left: 75% of the participants preferred the *Ribbon* interface over *1-Pillar* and *2-Pillar* when asked “Which training interface do you like best?” Right: The average results for hedonic and pragmatic quality and the 95% confidence intervals of the AttrakDiff2 questionnaire. The training interface had a significant effect on the hedonic quality

4.2 Raw NASA TLX

To measure users’ perceived workload we used the NASA Task Load Index (TLX) questionnaire (Hart and Staveland 1988). The NASA TLX is a multi-dimensional scale to obtain workload estimates from users while performing a task or immediately afterwards. We interpret the results of the unweighted, raw NASA TLX (RTLX) data. The RTLX is equivalent to the TLX but eliminating the weighting process (Hart 2006). The questionnaire consists of six subscales: mental (RTLX-MD), physical (RTLX-PD), and temporal demands (RTLX-TD), performance (RTLX-P), effort (RTLX-E), and frustration (RTLX-F).

Figure 3 (b) illustrates the results of the RTLX questionnaire. The results show that the *Ribbon* interface was more mental ($M = 38.75$, $SD = 27.44$) and physical demanding ($M = 48.33$, $SD = 24.38$) than the *1-Pillar* and the *2-Pillar* interface. The temporal effort (RTLX-TD) was highest for the *1-Pillar* interface ($M = 36.53$, $SD = 25.41$). On average, the exercise performance (RTLX-P) was rated best for the *2-Pillar* interface ($M = 38.89$, $SD = 21.89$). Using the *Ribbon* interface the participants had the impression of the lowest performance ($M = 31.39$, $SD = 27.45$). The results of the RTLX-E show that participants described their effort (mentally and physically) to accomplish the exercise goals lowest when using the *Ribbon* interface ($M = 43.06$, $SD = 24.07$). Additionally, the *Ribbon* interface caused less frustration ($M = 24.17$, $SD = 24.85$). Highest frustration was caused by the *1-Pillar* interface ($M = 32.78$, $SD = 25.86$).

One-way repeated measures ANOVAs yielded no significant effects of the *TI* for the variables RTLX-MD ($F(2, 70) = 0.89$, $p = 0.41$, $\eta^2 = 0.00$), RTLX-PD ($F(2, 70) = 0.39$, $p = 0.68$, $\eta^2 = 0.00$), RTLX-TD ($F(2, 70) = 0.60$, $p = 0.55$, $\eta^2 = 0.00$), RTLX-P ($F(2, 70) = 1.22$, $p = 0.30$, $\eta^2 = 0.02$), RTLX-E ($F(2, 70) = 0.48$, $p = 0.62$, $\eta^2 = 0.00$), and RTLX-F ($F(2, 70) = 1.60$, $p = 0.21$, $\eta^2 = 0.02$).

4.3 System Usability Scale and Interface Ranking

The participants evaluated each interface condition of the SMART-FIT system using the System Usability Scale (SUS). Ten items were rated on a 5-point Likert scale from 1 (disagree) to 5 (agree totally). The item scores were totaled to calculate an overall SUS score between 0 and 100, where higher scores indicate better usability (Brooke 1996). A mean SUS score above 72.75 reflects a good usability. A score larger than 85.58 refers to an excellent usability (Bangor, Kortum, and Miller 2008).

One participant forgot to fill in the SUS questionnaire for the *1-Pillar* condition. On average, all three interfaces were rated ‘good’ or better, i.e., a SUS score above 72.75. Participants perceived the usability of the *Ribbon* interface ($M = 81.14$, $SD = 19.17$) better than the *1-Pillar* ($M = 78.07$, $SD = 13.08$) and the *2-Pillar* interface ($M = 77.00$, $SD = 16.08$). A one-way repeated measures ANOVA showed no significant difference in mean SUS scores, $F(2, 68) = 0.86$, $p = 0.43$, $\eta^2 = 0.01$.

The result of the interface rankings is shown in Figure 4 (left). The ranking was analyzed by assigning scores from 3 to 1 in decreasing order from the highest ranked to the lowest ranked training interface version. 75 % of the participants ranked the *Ribbon* interface on first place. The participants rated the *Ribbon* interface best, following *1-Pillar* on 2nd place and *2-Pillar* on third place.

4.4 AttrakDiff2

To measure pragmatic and hedonic quality of the different *TIs* we used the German version of the AttrakDiff2 questionnaire (Hassenzahl, Koller and Burmester 2008). The questionnaire uses semantic differentials to assess the user experience of a product. From the AttrakDiff2 questionnaire we used the categories pragmatic quality (PQ) and hedonic quality (HQ). The questionnaire consisted of 21 items. The portfolio plot summarizes the results of the AttrakDiff2 questionnaire (see Figure 4, right). Participants rated the *Ribbon* interface best in terms of pragmatic quality ($M = 3.67$, $SD = 1.53$) and hedonic quality ($M = 4.67$, $SD = 1.34$). On average the *1-Pillar* received the lowest values for the hedonic quality. The *2-Pillar* interface got the lowest scores for the pragmatic quality ($M = 3.19$, $SD = 1.58$). A one-way repeated measures ANOVA shows a significant effect of the *TI* on the HQ, $F(2, 70) = 14.40$, $p < 0.01$, $\eta^2 = 0.12$. Bonferroni post hoc tests show a significant difference between *Ribbon* and *2-Pillar* ($M = 4.06$, $SD = 1.04$, $p < 0.01$) and *1-Pillar* ($M = 3.57$, $SD = 1.17$, $p < 0.01$), as well as between *2-Pillar* and *1-Pillar* ($p < 0.05$). The differences between the *TI* for the variable PQ were not significant, $F(2, 70) = 1.36$, $p = 0.26$, $\eta^2 = 0.02$.

4.5 Error Measurements

To evaluate the effect of the *TI* conditions on the performance we used different error measurements to compare the trajectory of the actual weight displacement with the trajectory of the target weight displacement. The following section will present the results of the calculated error measurements.

COR. The COR measures the linear correlation between two variables, i.e., the actual and the target weight displacement. +1 is a total positive correlation, 0 is no linear correlation. Mauchly's test indicate that the assumption of sphericity has been violated, $W = 0.59$, $p < 0.05$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = 0.71$). The results show that COR was significantly affected by the *TI*, $F(1.42, 49.64) = 3.76$, $p < 0.05$, $\eta = 0.01$. Bonferroni post hoc tests revealed a significant difference between COR for *Ribbon* ($M = 0.92$, $SD = 0.22$) and *2-Pillar* ($M = 0.88$, $SD = 0.22$). No other comparisons were significant.

MAE. The MAE measures the difference between the two continuous variables. It includes the comparison between the actual and target weight displacement. Mauchly's test for sphericity is significant, $W = 0.63$, $p < 0.05$. Thus, we corrected the degrees of freedom using Greenhouse-Geisser estimates ($\epsilon = 0.73$). A one-way repeated measures ANOVA results in a significant effect of *TI*, $F(1.46, 51.15) = 9.27$, $p < 0.01$, $\eta = 0.03$. The comparison of target and actual trajectory yielded the lowest MAE for the interface *Ribbon* ($M = 6.48$, $SD = 6.37$) followed by *1-Pillar* ($M = 7.87$, $SD = 4.66$) and *2-Pillar* ($M = 8.61$, $SD = 5.23$). Bonferroni post hoc tests show a significant difference between *Ribbon* and *2-Pillar* ($p < 0.001$). The differences between *1-Pillar* and *2-Pillar* ($p = 0.22$) respectively *Ribbon* ($p = 0.11$) were not significant.

RMSE. When using MAE to compare the trajectories one may neglect the impact of big but infrequent errors, i.e., large deviations from the target trajectory. To address this problem we calculated the RMSE. The lowest RMSE was computed for the *Ribbon* interface ($M = 8.30$, $SD = 8.64$). The RMSE for the *2-Pillar* interface was higher ($M = 11.24$, $SD = 6.67$) than for the *1-Pillar* interface ($M = 10.25$, $SD = 6.22$). Since Mauchly's test of sphericity was significant, $W = 0.71$, $p < 0.05$, we applied Greenhouse-Geisser estimates ($\epsilon = 0.77$) to correct the degrees of freedom. The one-way repeated measures ANOVA showed a significant difference for the dependent variable RMSE, $F(1.55, 54.14) = 7.89$, $p < 0.01$, $\eta = 0.03$. Again, Bonferroni post hoc tests show a significant difference between *Ribbon* and *2-Pillar* ($p < 0.001$). No other comparisons were significant (all $ps > 0.05$).

Russell's Errors. COR, MAE, and RMSE provide a more general measurement to compare target and actual trajectory. However, those measures do not account for a magnitude error (i.e. an error with regard to the ROM). Additionally, correctly but phase-shifted or asynchronously repetitions will result in a huge error. To tackle this problem, we used Russell's error measurements (Russell 1997). Russell developed a set of magnitude (RUS-M), phase (RUS-P), and comprehensive (RUS-C) error measures that address the

limitations of the previous error measurements. Following, we will describe the results for RUS-M and RUS-P using the different *TIs*.

On average, RUS-P is lowest for the *Ribbon* interface ($M = 0.0013$, $SD = 0.0018$) followed by *1-Pillar* ($M = 0.0022$, $SD = 0.0014$). *2-Pillar* has the highest value for RUS-P ($M = 0.0024$, $SD = 0.0015$). Mauchly's test indicate that the assumption of sphericity has been violated, $W = 0.82$, $p < 0.05$, therefore degrees of freedom were corrected using Huynh-Feld estimates of sphericity ($\epsilon = 0.89$). The results show that RUS-P was significantly affected by the *TI*, $F(1.78, 62.26) = 9.18$, $p < 0.05$, $\eta^2 = 0.03$. Bonferroni post hoc tests revealed a significant difference between RUS-P for *Ribbon* and *2-Pillar* ($p < 0.001$). No effect was found between *Ribbon* and *1-Pillar* ($p = 0.15$) and *1-Pillar* and *2-Pillar* ($p = 0.13$). These results suggest that the phase shift between target and actual trajectory was significantly smaller when participants used the *Ribbon* interface instead of the *2-Pillar* interface.

The lowest average RUS-M value is calculated for the *Ribbon* interface ($M = -0.0009$, $SD = 0.0024$). The magnitude error for the *1-Pillar* interface ($M = -0.0018$, $SD = 0.0016$) is slightly higher than for the *2-Pillar* interface ($M = -0.0016$, $SD = 0.0016$). A one-way repeated measures ANOVA using Huynh-Feld estimates ($\epsilon = 0.80$) shows a significant difference between the interface conditions, $F(1.61, 56.33) = 3.94$, $p < 0.05$, $\eta^2 = 0.04$. Post hoc comparisons using Bonferroni resulted in no significant differences.

5. DISCUSSION AND CONCLUSION

We used *Flow Experience* and *Worry* of the FSS to evaluate the effect of the training interface. The overall *Flow Experience* during the exercise was between 5 and 6 for all interfaces which is a typical value for engaging in sport activities (Rheinberg, Vollmeyer, and Manig 2005). No significant difference was found for the *Flow Experience*. However, participants were significantly more worried about doing the exercise correctly when using the *Ribbon* interface than using the *2-Pillar* interface. This is in line with the weak performance rating RTLX-P and the larger values for RTLX-MD and RTLX-PD.

The *Ribbon* interface yielded the highest average ratings for the categories RTLX-MD and RTLX-PD. On the one hand, participants perceived the *Ribbon* interface more physical and mental demanding and were less satisfied with their performance during the exercise. On the other hand, they have been less frustrated when using the *Ribbon* interface. This is also reflected in the good usability results and the ranking of the interfaces. As described above, the *Ribbon* interface provides a continuous feedback for all exercise parameters including the target displacement trajectory. Due to this participants might have focused continuously on the correct velocity respectively rotation angle when using the *Ribbon* interface. This might also explain the larger values for the flow experience. In contrast, using the *1-Pillar* and *2-Pillar* interface the user focused primarily on the correct ROM and only then on the correct velocity.

The AttrakDiff2 questionnaire evaluated the hedonic and pragmatic quality of the *TIs*. The pragmatic quality focuses on the usability of the product while the hedonic quality describes how innovative and valuable a product is. Overall, the *TIs* have only a neutral attractiveness and the 95%-CIs is quite large for the hedonic factor. This might be due to our non-homogeneous population in terms of age and gym habits. Future work will have to address this fact.

Although, participants rated the RTLX-P lowest when using the *Ribbon* interface, this fact is not reflected in the error measurements. Statistical analysis revealed significant effects of *TI* for all selected error measurements. Post hoc pairwise comparisons of the interface conditions gave significant lower errors when participants used the *Ribbon* than the *2-Pillar* interface. No significant differences were found between *Ribbon* and *1-Pillar*. *Ribbon* and *1-Pillar* interface combine the relevant exercise parameters – e.g. ROM and velocity – in a single visualization. When using the *2-Pillar* interface participants had to switch continuously the focus between two tasks: (1) align to target velocity in the left pillar and (2) check for the correct ROM in the right pillar. This dual-task problem might be the reason for significantly larger errors when using the *2-Pillar* interface. The errors in the magnitude (RUS-M) were not significantly different, when compared pairwise. Using the developed *Ribbon* interface instead of the *2-Pillar* interface would allow users to focus on the individual ROM and exercise velocity at the same time and to improve their exercise performance.

This paper presented the SMART-FIT system for strength exercise training. The system comprises a cable machine, notebook and a tablet PC. The tablet PC provides visual feedback during exercise training. Within this contribution we presented and evaluated the interface conditions *1-Pillar*, *2-Pillar*, and *Ribbon* to

support strength exercise training. We conducted a within-subject user study with 36 participants to evaluate the usability and performance of the developed interfaces. The results of the SUS and AttrakDiff2 questionnaire show a good usability of all interfaces. When asked to rank the interface, 75% of the participants preferred the *Ribbon* interface. Surprisingly, participants perceived their exercise performance (RTLX-P) better when using *2-Pillar* or *1-Pillar* interface. In contrast, the objective error measurements revealed that participants performed the exercise repetitions better using the *Ribbon* interface.

To date, we only evaluated the interface conditions for a simple exercise movement, i.e., the shoulder exercise. Additionally, we did not consider the physical and mental fatigue of users, since we only evaluated 10 exercise repetitions with a weight of 1 kg. Future work should focus on different strength and more complex exercise movements. This should include the personalized selection of the exercise weights to evaluate the different interfaces when muscles fatigue. What about the performance when the user reaches his maximal workload? Furthermore, complex exercises might lead to compensation movements. In our user study we noticed that participants sometimes lifted their elbow or moved their back while performing the exercise. These errors are not reflected in the error measurements. Additional movement sensors fastened to the user's upper arm might address these problems.

ACKNOWLEDGEMENT

We would like to thank all our participants volunteering in the user study. The research group wearHEALTH is funded by the German Federal Ministry of Education and Research (16SV7115).

REFERENCES

- Bangor, A.; Kortum, P. & Miller, J. A. (2008), 'The system usability scale (SUS): An empirical evaluation', *International Journal of Human-Computer Interaction* 24(6), 574--594.
- Brooke, J. (1996), *SUS - A quick and dirty usability scale*, CRC Press.
- Chittaro, L. & Sioni, R. (2014), 'Evaluating mobile apps for breathing training: The effectiveness of visualization', *Computers in Human Behavior* 40, 56--63.
- Eriksson, M.; Halvorsen, K. A. & Gullstrand, L. (2011), 'Immediate effect of visual and auditory feedback to control the running mechanics of well-trained athletes', *Journal of sports sciences* 29(3), 253--262.
- ERGO-FIT GmbH & Co. KG (2019), 'Smart fitness devices. www.ergo-fit.de.
- Hart, S. G. & Staveland, L. E. (1988), 'Development of NASA-TLX Results of Empirical and Theoretical Research, Human Mental Workload', *British Journal of Social Psychology*, 139--183.
- Hart, S. G. (2006), 'Nasa-Task Load Index (NASA-TLX); 20 Years Later', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50(9), 904-908.
- Hassenzahl, M.; Koller, F. & Burmester, M. (2008), 'Der User Experience (UX) auf der Spur: Zum Einsatz von www.attrakdiff.de, in Henning Brau; Sarah Diefenbach; Marc Hassenzahl; Franz Koller; Matthias Peissner & Kerstin Röse, ed., 'Tagungsband UP08', Fraunhofer Verlag, Stuttgart, pp. 78-82.
- Morris, D.; Tan, H.; Barbagli, F.; Chang, T. & Salisbury, K. (2007), 'Haptic Feedback Enhances Force Skill Learning, in 'Proceedings of the Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems', IEEE Computer Society, Washington, DC, USA, pp. 21--26.
- Rheinberg, F.; Vollmeyer, R. & Engeser, S. (2003), 'Die Erfassung des Flow-Erlebens, Hogrefe-Verlag, pp. 261 -- 279.
- Rheinberg, F.; Vollmeyer, R. & Manig, Y. (2005), 'DFG-Abschlussbericht zum Projekt Rh 14/11-1: Flow-Erleben: Untersuchungen zu einem populären, aber unterspezifizierten Konstrukt', Psycholog. Institut Universität Potsdam.
- Russell, D. M. (1997), 'Error Measures for Comparing Transient Data: Part I: Development of a Comprehensive Error Measure Part II: Error Measures Case Study, in 'Proceedings of the 68th Shock and Vibration Symposium', SAVIAC, Shock and Vibration Information Analysis Center.
- Sigrist, R.; Rauter, G.; Riener, R. & Wolf, P. (2013), 'Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review', *Psychonomic Bulletin & Review* 20(1), 21--53.
- Steffen, D.; Bleser, G.; Weber, M.; Stricker, D.; Fradet, L. & Marin, F. (2011), 'A Personalized Exercise Trainer for Elderly, in 'Pervasive Health Conference 2011. International ICST Conference on Pervasive Computing Technologies for Healthcare (Pervasive Health-2011), 5th, May 23-26, Dublin, Ireland', IEEE Xplore.