Comparing Modalities for Kinesiatric Exercise Instruction

Jan David Smeddinck
Digital Media Research Group
University of Bremen
Bibliothekstr. 1
D-28359 Bremen, Germany
smeddinck@tzi.de

Jens Voges
Digital Media Research Group
University of Bremen
jvoges@tzi.de

Marc Herrlich
Digital Media Research Group
University of Bremen
mh@tzi.de

Rainer Malaka
Digital Media Research Group
University of Bremen
malaka@tzi.de

The definitive version was published as: Smeddinck, J. D., Voges, J., Herrlich, M., & Malaka, R. (2014). Comparing Modalities for Kinesiatric Exercise Instruction. CHI '14 Extended Abstracts on Human Factors in Computing Systems, 2377–2382. https://doi.org/10.1145/2559206.2581367

Abstract
We present an experimental comparison of three kinesiatric exercise instruction modalities: a live human instructor (*human*), recorded video (*video*) and a virtual figure displayed next to the representation of the users' approximate skeleton (*interactive*). The results regarding user experience, preferences, and exercise accuracy indicate a preference for the *human* instructor across measures. A disparity exists between exercise accuracy and perceived ease of understanding when comparing the *video* with the *interactive* modality. Perception measures indicate a slight preference for the video modality, whilst performance data shows a significantly higher accuracy in the interactive condition. Our findings support the further investigation of digital interfaces to support physical therapy and rehabilitation as a cost-effective and potentially more efficiently customizable addition to traditional exercise instruction forms.

Author Keywords
Embodied interaction; natural user interface; games; accuracy; motion-based input; feedback

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.
K.8.0 [Personal Computing]: General - Games.

Instruction Modalities
(a) – *human*; the gold standard in practical PT/R for exercises that do not require additional tools or machinery. A sole instructor was trained to consistently deliver the exercise set. Each study session was accompanied by a pre-recorded audio voice-over to assure temporal stability. The track was used across all conditions (a, b & c).
(b) – *video*; a commonly employed modality for the at-home delivery of kinesiatric exercise instructions. In this study it serves as a baseline. The video was created by recording the trained instructor.
(c) – *interactive*; a bare-bones implementation of an "instructions and feedback representation" scheme that is achievable with consumer grade FBMTS and matching interactive applications. A comparison of the other modalities with the given tool, lacking all decorations and further game mechanics, is cogent to isolate the impact of the representation method from the potential influence of other elements of a more complex "real" game or a more information rich application interface. The skeletal data for driving the virtual reality (VR) character was recorded in the same take as the instruction video.

Introduction
A growing body of research investigates full-body motion-tracking sensors (FBMTS) and motion-based games (MBGs) as elements of applications that are intended to support physical therapy or rehabilitation (PT/R) by offering motivation, guidance and/or progress information. Recent studies have underlined the potential of such applications with a number of target groups (e.g. [3,6]) and others have evaluated the tracking accuracy of consumer grade FBMTS, such as the Kinect [1]. While professional solutions can theoretically deliver a higher accuracy, they often have their own sets of limitations (e.g. increased cost, challenges in calibration, intrusivity, maintenance, real-time usage) [1] and the tracking accuracy of the consumer hardware is reasonable for many use-cases. FBMTS-driven interactive applications have been studied in comparison to traditional exercise guidance by instruction sheets [8]. However, to the best of our knowledge, the question: *"How does a full-body motion-based interactive solution compare to more traditional kinesiatric exercise instruction modalities when conveying actual physical exercise sets, in terms of the resulting user experience and the exercise performance?"* remains unanswered. We thus present an experimental comparison of three alternative modalities: (a) live human-to-human instruction [*human*], (b) instruction by video, with a recorded human instructor [*video*] and (c) instruction via an interactive tool [*interactive*] that shows a stick-figure animation of pre-recorded exercises together with a stick-figure representation of the recipient.

Related Work
Existing work on the potential of FBMTS in MBGs for health has focused on the Kinect [5] and similar technologies, as well as other devices, such as digital balance boards [4], pointing devices [2], etc. Since no-contact full-body motion-tracking devices are arguably most flexible and can support a large number of different exercises and use-cases, the question arises, how *interactive* exercise instruction compares to other modalities. Uzor et al. [7] argue that, next to the scenario of augmenting live instructor-to-recipient/group sessions, the most established method that this new technology competes with are instructional booklets, or videos, since those do not require the presence of a therapist. They found that the latter modalities may have a negative influence on the adherence to home rehabilitation programs (e.g. skipped repetitions / exercises, movements faster than intended, passive nature, no progress display). A follow-up study compared instructional booklets and more interactive forms of exercise instruction [8] offering either passive feedback (a guiding mannequin only), or real-time feedback (a mannequin reproducing the movements of the user). Participants reportedly paid more attention to the guiding mannequin, but experienced feedback provided by the real-time mannequin to be helpful. Interactive methods were preferred over the instructional booklets and the authors conclude that they support the participants in performing the exercises at the correct speed [8]. However, comparative studies with further alternative instruction modalities are needed to provide insights on the exact benefits that can be expected - and potential trade-offs that may be involved, when choosing between the available modalities.

Experiment Design & Research Methods
We therefore investigate whether exercises are still properly perceived when they are presented in a plain

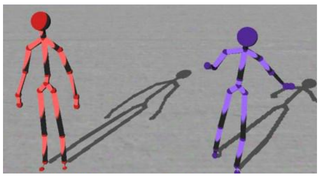

Figure 1: A screenshot of the interface for the *interactive* exercise instruction modality (l: instructor, r: participant). Shadows added for better perceptibility of orientation. Elevated camera in instruction demonstration phase. Motion data was captured with a Kinect device.

Participants

All participants (10 male, 10 female, average age M=50.35 years, SD=20.72, MIN/MAX=15/86) participated on a voluntary basis. 85% of the participants reported regular computer use and 16 participants stated that they play digital games less than once a month or never. 60% reported to be performing physical exercises of arbitrary kind at least once a week. None of the participants stated that they perform more than eight hours of physical activity per week. Eight participants stated to have done physical exercise programs which involved choreographies.

virtual environment, a generic setting in which a wide range of exercise patterns can be realized, even specifically for individual patients, and which also represents the technological basis for a wide range of applications, and how this modality compares to *video* or *human* instructions. The *Instruction Modalities* text box provides further information on each modality.

The study follows a within-subjects design with the instruction modality as the independent variable and the dependent variables of user experience (UX), preference and performance accuracy, which we aim to capture via triangulation by combining multiple measures. Affect was measured before the experiment and after each condition by using a self-assessment manikin (SAM). A Borg rating of perceived exertion (RPE) was also collected following each condition. Established UX instruments do not match all interaction schemes in this study. We therefore collected responses to three Likert-scale items that are specific to the research at hand, following each condition: the participants' perceived *enjoyment*, their ease of *understanding* the exercise instructions, and their self-evaluated performance *accuracy* (cf. Table 2). After completing all trials, participants ranked the modalities according to the same three aspects and were asked to provide reasons for the ranking decisions.

The exercises employed in the study were selected from an instruction DVD with exercise routines for Parkinson's disease patients[1]. Five exercises that involve either upper and/or lower body movements and can be performed while standing upright without requiring any equipment were selected. Each exercise was first introduced verbally whilst being demonstrated and the recipients were then asked to perform ten repetitions of the exercise together with the instructor, while a verbal count-down provided cues for improved alignment and clarity about the progress. Each exercise session lasted 2 minutes and 17 seconds. Conditions were presented in Latin square randomized order. The study was conducted with healthy convenient subjects to establish first insights sparing the complexity of factors introduced by target groups with special needs (cf. to the *Participants* text box for further details).

[1] "Obenauf bei Parkinson" (2008), Desitin Arzneimittel GmbH

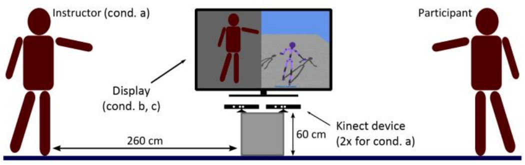

Figure 2: Experiment setup (mixed display of all conditions).

The experiment setup differed slightly between the *human* condition and the *video*, as well as the *interactive* condition (see Figure 2). Exercises were performed in a mirrored fashion. The method for measuring the distance between posture data sets between conditions and participants is summarized in the *Comparison of Exercise Movement Data* text box.

Results

All interval-type responses were analyzed with one-way repeated-measures ANOVAs (α=0.05) with Mauchly sphericity tests. Post-hoc pairwise t-test comparisons are performed at Šidák corrected (via SPSS) significance levels upon a positive indication of a null hypothesis violation in the ANOVAs.

Comparison of Exercise Movement Data

Strategy applied to compensate for different body proportions between participants:

1. Predetermine relevant data segments based on audio-cues.
2. Isolate repetitions 3 to 8 (out of 10)
3. Remove noisy posture data (1 participant removed of 20).
4. Mirror recipient skeleton (*human* condition only).
5. Translate both skeletons so that the central hip joint resides in the origin.
6. Scale the recipient's skeleton so that the central hip bone to central shoulder bone distance matches that of the instructor's skeleton.
7. Calculate local error based on sum-of-squares distances (calculated with the Frobenius norm) between the gold-standard hierarchical joint orientations of the instructor and those of the recipient, for a preselected subset of joints that are relevant to each exercise.
8. Provide a grand mean of the sum of all offsets over all frames for each exercise.

The RPE responses display nearly identical means after all conditions (*human*: M=9.50, SD=2.35; *video*: M=9.40, SD=2.28; *interactive*: M=9.40, SD=2.30) with no significant differences between conditions. Overall levels indicated "very light" exertion (corresponds to heart-beat rates of ~ 90 to 100 BPM).

An overview of the 7pt. scale SAM affect responses is presented in Table 1. A statistically significant difference was found for the valence measure ($F(3,57)=3.473$, $p=.022$, $\eta^2=.155$). The post-hoc pairwise comparison showed a significant difference between the *before* and the *human* condition ($p=.045$). Arousal responses were corrected with Greenhouse-Geisser due to sphericity violation ($\chi^2(5)=11.109$, $p=0.05$). Results indicate significant differences within-subjects ($F(2.103, 39.955)=9.879$, $p<.001$, $\eta^2=.342$), with post-hoc pairwise-comparisons showing significant differences between *before* and each of the conditions (all $p<.001$), but not between any of the conditions. The ANOVA on the dominance responses did not indicate any significant difference.

The participants' responses to the use-case specific 7pt. Likert-scale items, which were presented with edge labels ranging from *"strongly disagree"* to *"strongly agree"* and were concerned with perceived *enjoyment*, *understanding* and *performance accuracy*, are summarized in Table 2. The ANOVA for *enjoyment* did not indicate significant differences. However, together with the rather strong mean agreement, the data showed a notable distribution, with 45%/40% of participants indicating *"strongly agree"* following the *human/interactive* condition respectively, whereas only 30% indicated *"strongly agree"* following the *video* condition. The response analysis on *understanding* did indicate the existence of a significant difference ($F(2,38)=7.954$, $p=.001$). Post-hoc tests point at the conditions *human* and *interactive* ($p=.01$); the instruction modality employing a human instructor was perceived to be more easy to understand. Lastly, the ANOVA of the perceived *imitation accuracy* also shows a significant effect ($F(2,38)=5.093$, $p=.011$), which can be explained by a significant difference between the *human* and the *video* condition ($p=.044$) where the former received a higher mean agreement.

An overview of the ranking responses is provided in Figure 3. A Friedman test was applied to each ranking (α=.05) and post-hoc Šidák corrected Wilcoxon Signed-Rank tests (significance level here: p<.033) were employed to qualify differences between conditions. The *enjoyment* ranking did not show a significant difference, but the *understanding* ranking ($\chi^2(2)=9.1$, p=.011) and the *accuracy* ranking ($\chi^2(2)=6.7$, p=.035) did. The difference in understanding and accuracy was significant between the pair of the *human* and the *interaction* condition (p=.018 and p=.025 respectively). The distribution of 1st, 2nd and 3rd places shows that *video* is most commonly in 2nd place, whereas for some users *interactive* did take first place. With regard to ease of understanding, most of the participants ranked the *human* condition first. Reasons were that *"the real body image is more understandable"*[2], a possibility of a *"direct comparison"* or the larger size presentation. Participants who ranked the *interactive* condition first said *"the mannequin was the easiest to perceive"* or that it was helpful to see oneself. Participants who ranked the interactive condition first for enjoyment mentioned reasoned: *"it was most interesting"*,

[2] All participant quotes were translated from German.

Table 1: SAM results. v=valence, a=arousal, d=dominance.

Before	A (human)	B (video)	C (interactive)
vM=5.25, vSD=1.02	vM=5.80, vSD=0.7	vM=5.80, vSD=0.7	vM=5.35, vSD=1.09
aM=3.45, aSD=1.28	aM=4.30, aSD=1.38	aM=4.45, aSD=1.47	aM=4.55, aSD=1.57
dM=4.65, dSD=0.99	dM=4.90, dSD=1.29	dM=5.15, dSD=1.09	dM=4.80, dSD=1.28

Table 2: An overview of the use-case specific Likert-scale items presented after each condition, together with the response means and standard deviations

Statement / Trial	A (human)	B (video)	C (interactive)
"I enjoyed doing this set of exercises."	M=5.95, SD=1.19	M=5.65, SD=1.14	M=5.65, SD= 1.46
"The shown exercises were easy to understand."	M **6.60**, SD=0.68	M=6.15, SD=0.81	M **5.60**, SD=1.43
"I could imitate the exercises accurately."	M **6.05**, SD=1.0	M=5.30, SD=1.17	M **5.25**, SD=1.37

"because I could see my own movements", or *"the animation can be comprehended best"*. Reasons stated for placing the human condition first for enjoyment were: *"doing the exercises together is fun"*, *"more intuitive […] better proportions"*, *"eye contact"*, or *"interaction with another human being"*.

Performance Analysis

Since distance metric results between skeletal movement data are not well comparable in absolutes, the sums of mean rotation differences for each exercise in every condition were normalized per participant against the condition in which that participant performed with the largest difference in comparison to the instructor posture data. The measure can be summarized as the *normalized mean squared (relevant hierarchical joint) rotation distance* (NMSRD). The resulting relative differences between conditions of all participants (cf. Figure 4) were subjected to a one-way *rm*ANOVA. Results indicated a significant effect on this measure ($F(2,36)=39.608$, $p<.001$, $\eta^2=.688$). Post-hoc pairwise comparisons showed significant differences between all trials ([b|a, b|c]: $p<.001$, [a|c]: $p=.002$). The difference in performance compared to the original instructor's posture data was significantly higher in the *interactive* condition than in the *human* condition and it was again significantly higher in the *video* condition than in both aforementioned conditions.

Discussion

The results suggest that the human instructor modality is the most preferred one and leads to the most accurate performance. The RPE results expectedly indicate consistently low levels of physical exertion across conditions, suggesting that all modalities can be expected to deliver similar levels of exertion. We link the slightly increased arousal compared to pre-trial levels across conditions to this result. The use-case specific perception items and the participants' statements indicate that the instructions in the *interactive* modality appear (to some) to be difficult to understand. This condition also did not lead to significantly increased valence compared to the pre-trial levels. However, the ratings for understanding are high and the high enjoyment ratings did not suffer in comparison to the other modalities. The performance accuracy was significantly higher in the *interactive* condition than in the *video* condition, even though the perceived accuracy does not mirror this objective finding. The varying indications with regard to user experience and preference between the *video* and the *interactive* modality warrant further consideration. The low-fidelity visual representation may appear alien to many users who are not accustomed to virtual reality displays. This may also explain the lower ratings for ease of *understanding* with this modality. The participants' perception of accuracy did correspond to objective accuracy differences between the human condition and the other conditions, but not in the difference between the video and the interactive condition. This serves as a reminder of the importance of considering both subjective perception and objective performance.

The objective offset in performance accuracy between the *human* instructor modality and the two alternatives indicates that constantly identical representations of the exercise instruction delivery, which was present both in the *video* and the *interactive* modality, does not seem to lead to increased performance accuracy compared to the gold-standard. The fact that the *interactive* modality did beat the baseline in terms of objective performance accuracy may be attributed to the presence of a virtual representation of the exercise recipient. The strong finding that performance accuracy was best in the *human*

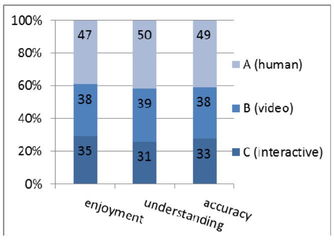

Figure 3: Combined ranking scores per condition for each of the three ranking items.

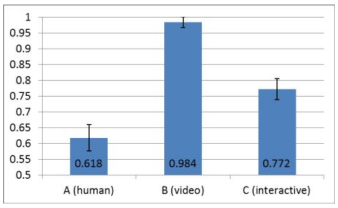

Figure 4: The mean NMSRD (normalized mean squared rotation distance) across all participants for each condition (error bars indicate standard error).

modality may be due to the comparatively rich display of live human performance, which was apparently lost in the video condition. Social factors may motivate participants to act in a more encouraged manner; however, future work is needed to further differentiate this result.

Conclusion

Considering all findings, we argue that motion-based technology enabled applications and games for the support of PT/R are worth-wile to pursue as a matter of research and development. While the perceptive measures in this study show that a bare-bones implementation of such an interactive kinesiatric exercise instruction modality does not clearly beat existing modalities in all aspects, it can result in higher performance accuracy than established alternatives and is usable and enjoyable even without making use of the full potential of digital media and games. While it is important to understand the effect of the general modality separately from other factors, we now have reason to assume that the positive indications in related work are to notable extends due to positive effects of decorations, additional functionalities, information and potential reward or game mechanics. Future work is needed to separate the impact of interpretability of a photorealistic / real human, consistent reproduction of instructions and the role of constant feedback in more detail.

Acknowledgments

This work was partially funded by the *Klaus Tschira Foundation (KTS)* through the graduate school *Advances in Digital Media*.

References

[1] Gabel, M., Gilad-Bachrach, R., Renshaw, E., and Schuster, A. Full body gait analysis with Kinect. *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (2012), 1964–1967.

[2] Gerling, K.M., Dergousoff, K.K., and Mandryk, R.L. Is Movement Better? Comparing Sedentary and Motion-Based Game Controls for Older Adults. *Graphics Interface 2013*, (2013), 133–140.

[3] Gerling, K.M., Livingston, I.J., Nacke, L.E., and Mandryk, R.L. Full-Body Motion-Based Game Interaction for Older Adults. *CHI'12: Proceedings of the 30th international conference on Human factors in computing systems*, (2012).

[4] Gerling, K.M., Schild, J., and Masuch, M. Exergame design for elderly users: the case study of SilverBalance. *Proceedings of the 7th International Conference on Advances in Computer Entertainment Technology*, (2010), 66–69.

[5] Lange, B., Koenig, S., McConnell, E., et al. Interactive game-based rehabilitation using the Microsoft Kinect. *Virtual Reality (VR), 2012 IEEE*, (2012), 171 –172.

[6] Smeddinck, J., Siegel, S., and Herrlich, M. Adaptive Difficulty in Exergames for Parkinson's disease Patients. *Proceedings of Graphics Interface 2013*, (2013).

[7] Uzor, S., Baillie, L., and Skelton, D. Senior designers: empowering seniors to design enjoyable falls rehabilitation tools. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2012), 1179–1188.

[8] Uzor, S. and Baillie, L. Exploring & designing tools to enhance falls rehabilitation in the home. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2013), 1233–1242.