Towards an Evaluation Framework: Implicit Evaluation of Sense of Agency in a Creative Continuous Control Task

Axel Hoesl
LMU Munich
Munich, Germany
axel.hoesl@ifi.lmu.de

Florian Lehmann
LMU Munich
Munich, Germany
lehmannf@cip.ifi.lmu.de

Phuong Anh Vu
LMU Munich
Munich, Germany
vuphu@cip.ifi.lmu.de

Andreas Butz
LMU Munich
Munich, Germany
andreas.butz@ifi.lmu.de

Christina Rosenmöller
LMU Munich
Munich, Germany
rosenmoeller@cip.ifi.lmu.de

 CHI'17 Extended Abstracts, May 06-11, 2017, Denver, CO, USA
ACM 978-1-4503-4656-6/17/05. http://dx.doi.org/10.1145/3027063.3053072

Abstract
Reducing workload in a UI while still letting users feel in control is not trivial. This workload/control tradeoff is described in the literature and deserves attention in design practice. However, there is no evaluation framework for it supporting both explicit and implicit measurement, mainly because measuring sense of control implicitly is just difficult. A recently proposed implicit evaluation methodology – measuring *sense of agency* via *interval estimation* – seems promising and calls for further investigation. We studied its feasibility in a continuous control task – cinematographic camera motion – and compared a multi-touch software joystick to a mid-air gesture UI (N=8). Data was collected both explicitly and implicitly. Our results suggest that the mid-air gesture design does not increase the sense of agency. Both methodologies yielded similar results but the implicit one was more sensitive and the combination of both led to convincing overall results.

Author Keywords
User interface; sense of agency; sense of control; implicit measures; interval estimation; intentional binding.

ACM Classification Keywords
H.5.2 [User Interfaces]: User-centered design, Evaluation/Methodology, Input devices and strategies.

Sense of Agency

Sense of Control is a recurring scheme with many facets in HCI. A long thread of research is concerned with identifying variables [10], developing design guidelines [17], studying systems [16, 3, 13] or adapting automation levels [1, 20]. It is ubiquitous in hardware [20], graphical [20] or post-WIMP [15] user interfaces as well as in virtual reality [4], semi-automated systems [1, 20] and robotics [22].

Systems nowadays use automation, UI adaptation or novel input modalities, and in consequence, their design and behavior increasingly influences not only the users' factual level of control but also their perception of it. Dominance effects can even lead to counter-intuitive situations in which less factual control is perceived as being more in control [20]. As pointed out in [14], delegating functionality to automation leads to an inevitable tradeoff between workload reduction (wanted) and unpredictability of the results (unwanted). Experiencing decreased control due to unpredictable results is generally unwanted and even more so in creative tasks. In this domain, technology is rather used as an extension of human capabilities and as a means of active personal expression than for the delegation of daily chores. Thus, unpredictable results let users reject a system more easily.

To avoid this pitfall, design alternatives must be evaluated thoroughly. Contributing to the existing tool palette, a new strand of psychological research examined human perception of control, summoned under the term *Sense of Agency* (SOA) [19, 1, 12] in recent years. It proposes a way of implicitly measuring this experience on a non-conceptual level, letting designers gather insights beyond self-reported data. The described methodology was already adopted in HCI research, but so far only few publications report on it. Apparently only *discrete* control operations were chosen as study objects [1, 3, 13]. In our work, we apply this approach to user interfaces for *continuous* control of camera motion.

Goal and Contribution

Reducing workload while maintaining a sense of control is a frequent goal of UI designs for semi-automated or adaptive systems. While this applies in general, it is especially important when designing tools for creative tasks. To evaluate both aspects in a coordinated way, an elaborate evaluation framework is needed. So far, the workload/control tradeoff can only be estimated on an explicit level by using specific questionnaires. Questionnaires for estimating personality traits and user perception have been proposed [21, 10, 5, 7] and are commonly used. Yet with questionnaires, data can only be gathered on a conceptual level. In addition, they interrupt a task or can only be rendered at the end of a trial.

On an implicit level however, data can be measured continuously and without revealing the underlying true intent of the measurement. For the introduced workload/control tradeoff [14], currently only workload measurements can be taken continuously during a task, e.g., using a Detection Response Tasks (DRT) [2]. Measuring sense of agency via *interval estimation* [19, 1, 12] could help to fill this gap. It is described as a more sensitive tool, but still it needs further exploration, a deeper understanding and validation in HCI in addition to the conducted fundamental research.

We actually propose that when integrated into a DRT, workload and sense of control can both be determined continuously during the same task and existing questionnaires can be used after the trial for further details. This creates a rather elaborate framework to better understand the effects of interface and system designs on both a non-conceptual and a conceptual level. The study we present constitutes a first step towards such a framework, by validating the implicit measurement tool with continuous control.

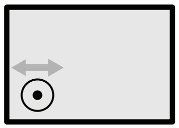


Figure 1: Tablet with software joystick interface.

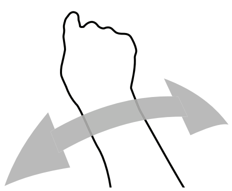


Figure 2: Mid-Air Gesture user interface.

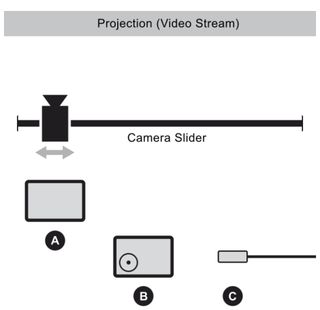


Figure 3: Study Apparatus with tablet displaying the video stream (A), tablet with software joystick (B) and Leap motion controller (C).

Measuring Sense of Agency Implicitly

In the limited space of this paper we will only briefly summarize the preceding work. An in-depth introduction to Sense of Agency as a concept and implicit measuring methodologies was provided by Berberian et al. in [1] and by Limerick and colleagues in [12]. One established evaluation technique for measuring sense of agency in an implicit way is *interval estimation*. It exploits distortion effects on the human perception of time during preconscious temporal binding processes. These distortions can be determined by sequentially presenting two stimuli to study participants and asking them to estimate the time interval in between. In general, the intervals to be estimated can range between ~150ms and ~1500ms before the effects break down. Persons who experience an increased sense of agency, experience a certain interval shorter than it actually is due to the distorting effects. They also report shorter estimates than persons who do not experience a sense of agency. Such distortions can be caused by an increased physical activity as cues necessary for the emerging of this experience are provided by internal processes serving motor control [8].

These suggestions and methodologies have been applied in the field of HCI before. In [3] a skin input prototype was compared to a traditional keyboard and mouse. The authors used implicit measurements and concluded that using skin input provided an increased sense of agency. Similarly, a speech recognition interface was compared to keyboard input in [13]. Here, the authors concluded from implicitly collected data that speech recognition did not lead to an increased sense of agency when compared to a traditional keyboard. Beyond the descriptions of results, the cited work provides insights into the study procedures. Our study conditions were inspired by the notion that increased physical activity can have effects on sense of agency and our procedure was modeled after studies of the presented literature.

Study using Implicit Measurements

Motivated by the findings above, we implemented and compared two interaction styles. As the literature indicates, physical activity in the control of a system can lead to an increased sense of agency (on non-conceptual level). We wanted to know whether this effect can be exploited for remote camera control by using a *mid-air gesture* user interface. In addition, we also wanted to test the feasibility of implicit measures in a continuous control task. We therefore compared this interaction style to an established technique in the domain, the *multi-touch software joystick*. This condition served as the baseline for remote control. In all conditions, data was collected implicitly via interval estimation as well as explicitly with questionnaires.

User Interfaces

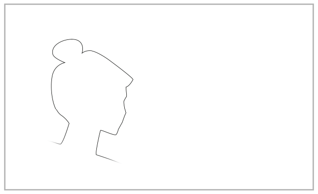
For the *multi-touch* condition we implemented a state of the art software joystick on an off-the-shelf Android tablet. The software joystick's horizontal position controlled the direction and speed of a connected camera motion control device. Such an interface is often found in physical camera control (e.g., for drones) or in virtual camera control for games. As a diminished sense of precision can be observed with multi-touch interfaces in general and as the touch-screen of a mobile device is inevitably occluded by interface elements, mid-air gestures could help to address both issues and also increase physical activity in controlling camera motion. The mid-air gesture condition was implemented using a Leap Motion for controlling the direction and speed via the pitch of the hand and forearm.

Participants

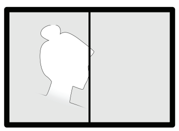
We recruited 8 participants (4 female) between 25 and 34 years with a median age of 26.5. To avoid interference of handedness in the mid-air gesture condition, we only invited right-handed people. All participants had normal or

Procedure

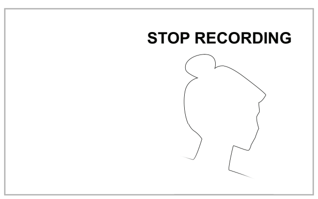
1. Video projection starts



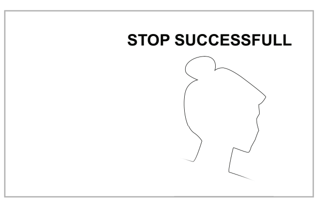
2. Video stream on tablet



3. First Cue



4. Delayed Second Cue



5. Interval Estimation

corrected to normal vision and were inexperienced in professional camera work. Two participants reported that they had used a gesture control interface before.

Study Design and Task
We designed a within-subjects study in a controlled laboratory environment. The two user interface conditions were presented in counter-balanced order to avoid learning effects. During the study a portrait photograph was projected onto a wall and filmed by a camera on a motion control system. The photograph was animated to move from left to right and participants were asked to follow it with the camera while framing it in the center of the screen.

Apparatus
The setup mimicked a production studio with a moving actor and remote controlled tool environment. For the display of the moving image in our framing task, we used a short distance wide angle projector. A portrait photograph moving horizontally from left to right was projected onto a wall in front of the participants. The projection was captured with a DSLR and transmitted wirelessly to a tablet to provide visual feedback. To support the framing goal of positioning the photograph in the center of the image-frame continuously while it moves, a line was displayed at the horizontal center of the video stream. The software was implemented in Java 8 running Processing 3.

Procedure
Our procedure was modeled after the study designs and procedures reported in [3, 13, 20] and adapted to our simple camera motion task. After welcoming the participants, they were informed about the study and asked to sign a declaration of intent. Having declared consent, the participants were presented the first user interface condition. Before the start of each trial, a short training phase (< 5 min) was given for each interface and the details of the task were explained. Following each interface training phase, also an interval estimation training phase was conducted: Participants were given two sequential audio stimuli and asked to guess the amount of time between them. After each guess the correct answer was revealed.

After this preparation phase, the assigned trial was started by the experimenter. The participants used the assigned user interface to steer the motion control system in order to follow the moving image and frame it in the center. After filming for 7 to 13 seconds, and after an additional random interval an audio-visual cue appeared, signaling to stop the task. With an action-effect delay of 500 ms, 800 ms or 1500 ms a second cue was presented signaling that the recording successfully stopped. The exact intervals were taken from related work [19, 1, 3].

Participants were then asked to estimate the interval (*implicit* measurement). Each task and measurement was repeated ten times for each condition. After finishing the tasks with one user interface, questionnaires on the sense of control (*explicit* measurement via SCS [7]), workload (Raw TLX [9]) and demographics were handed out. After filling out the questionnaires for all interfaces, the participants were debriefed and thanked for their participation.

Results
For the data collected on Interval Estimation Errors (IEE) and workload the Shapiro-Wilk Test was non-significant, thus normality can be assumed. Still, negative covariance could lead to a loss of statistical power for paired samples. We thus determined the covariances and found that all data sets (including SCS) have a positive correlation between both conditions. As an overestimated effect size could occur, due to a correlation between related samples we corrected effect size according to [11].

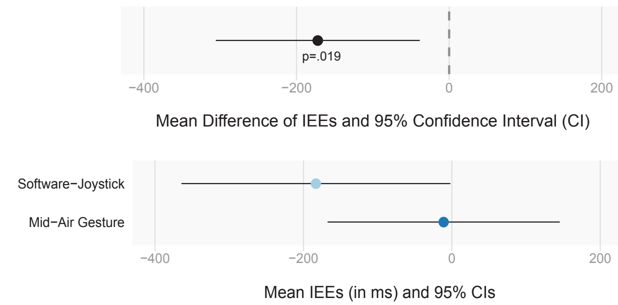


Figure 4: *Mean difference* of Interval Estimation Errors (IEE, top) and *mean interval estimation errors* (bottom).

Interval Estimation Errors
Assuming normality, we conducted a repeated measures ANOVA to compare IEEs in the tested conditions. There was a significant difference in the scores for *software-joystick* (M=-183.19, SD=216.75, CI[-364.40, -1.99]) and *mid-air gesture* (M=-11.04, SD=187.15, CI[-167.5, 145.42]); F(1,7)=9.28, p=.019, η^2=.40, ω_p^2=.48, 74%, CI[-305.82, -38.49]. These results suggest that the interaction style had an effect on the participants' sense of agency. More specifically, our results suggest that it decreases for the mid-air gesture condition.

Sense of Control Scale
Based on the work of Dong et al. [7] we collected self-reported data on the participants' sense of control via a 6-item rating scale. To compare the data collected in both conditions we ran a Wilcoxon-Signed-Rank test and found a significant difference in the scores for *software-joystick* (Mdn=4) and *mid-air gesture* (Mdn=2.5) conditions; Z=-1.98, p=.047, η^2=.25. This suggests that sense of control decreases for the mid-air gesture design.

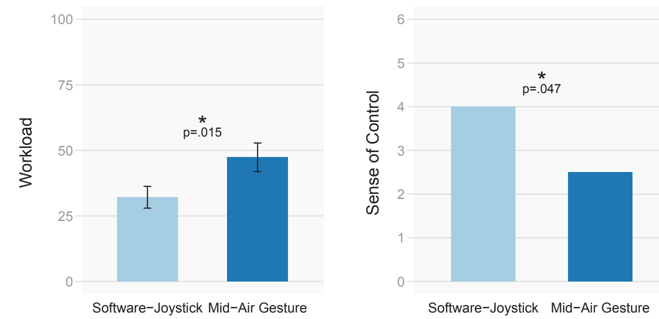


Figure 5: The explicit workload (left, p=.015) and sense of control (right, p=.047) scores differ significantly.

Raw Task-Load Index
We further measured workload via the Raw TLX [9] and averaged the collected items into a score ranging between 0 and 100. For both index scores, we conducted a repeated measures ANOVA. There was a significant difference in the scores for *software-joystick* (M=-32.13, SD=11.74, CI[22.31, 41.94]) and *mid-air gesture* (M=47.38, SD=15.4, CI[34.5, 60.25]) conditions; F(1,7)=10.31, p=.015, η^2 =.43, ω_p^2=.51, 79%, CI[4.02, 26.48]. These results suggest that interaction style has an effect on the perceived workload. In our case, workload increases with the mid-air gesture.

Preference for the Software Joystick
Compared to a state of the art software joystick, mid-air gestures as an interaction style for continuous control could not increase the perceived sense of agency. Thus, we have to discard this design choice when trying to increase sense of control. This was apparent from the explicit and the implicit data alike. However, the p-value in the explicit data analysis (p=.047) is based on a single 6-item rating scale and barely below the alpha level of 0.05. As simulation studies indicate, p-values between 0.04 and 0.05 are rather

unlikely and more surprising than convincing evidence [18]. So in consequence, to rely only on this measurement, would not necessarily provide strong support for our conclusion. However, we can take the implicitly collected data into account as well. The analysis of Interval Estimation Errors supports our prior conclusion of a measurable and actually significant difference. Consequently, we are more convinced of our interpretation due to the multiple measurements given the observed data. This is also supported by the workload measurement. It suggests that the mid-air gestures lead to an increased perceived workload despite the rather simple task.

Of course, these results are not a general verdict against mid-air gestures. They might still be suitable for discrete operations, such as starting, pausing or stopping camera motion or triggering automated tasks (e.g., moving closer towards the recorded subject). Although gestures could not outperform a software joystick in terms of sense of control, they still provide the benefit of not occluding the screen and allow triggering and coarse adjustments in an eyes-free way, e.g., when recording oneself during sports activities.

Evaluation Framework

Prior work suggests that implicit measurements are more sensitive evaluation tools and more likely to identify significant differences than explicit ones. This could also be observed in our study considering the smaller effect size η^2 of .25 for the explicit compared to .4 for the implicit data. So, should we substitute explicit with implicit measurements altogether? Working towards an elaborate evaluation framework, we propose to take both explicit and implicit measurements. On the one side, explicit measurements sometimes might be too coarse to identify all occurring effects (resulting in a Type 2 error). On the other side, as implicit measurements are more sensitive, they might also be more prone to false-positives (Type 1 errors). In combination however, the results can become more stable than either of the parts.

With implicit measurements alone, it is hard to guarantee, that an experiment is exactly measuring a specific phenomenon free from external interference. Thus, we recommend to additionally collect explicit data. For sense of agency, the question of whether both methodologies actually measure the same, has been raised by [6], concluding that they could tap into different processes of self-attribution of agency and control. The same holds for workload. This needs to be considered in the further necessary development steps towards an integrated evaluation framework for the workload/control trade-off.

Conclusion

Our goal was to compare two interaction styles for camera motion control and to explore the feasibility of implicit measurements of sense of agency in such a continuous control task. We compared a software joystick to a mid-air gesture interface for camera motion control in a user study (N=8). Our results indicate that despite an intrinsic increase in physical activity, the mid-air gesture design led to a diminished sense of agency. We collected both implicit measurements via interval estimation and explicit measurements with questionnaires. Both evaluation approaches showed significant differences between the conditions. However, the implicit measurement was more sensitive and provided additional necessary evidence for a convincing conclusion. Since implicit and explicit measurements led to similar conclusions and both types of measurements could still tap into different self-attribution processes, we suggest to use both in a more elaborate evaluation framework.

REFERENCES

1. Bruno Berberian, Jean-Christophe Sarrazin, Patrick Le Blaye, and Patrick Haggard. 2012. Automation Technology and Sense of Control: A Window on Human Agency. *PLoS One* 7, 3 (2012), e34075. DOI: **http://dx.doi.org/10.1371/journal.pone.0034075**

2. Antonia S. Conti, Carsten Dlugosch, and Klaus Bengler. 2012. Detection Response Tasks: How Do Different Settings Compare?. In *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '12)*. ACM, Portsmouth, New Hampshire, 257–260. DOI: **http://dx.doi.org/10.1145/2390256.2390298**

3. David Coyle, James Moore, Per Ola Kristensson, Paul Fletcher, and Alan Blackwell. 2012. I Did That! Measuring Users' Experience of Agency in Their Own Actions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'12)*. ACM, New York, NY, USA, 2025–2034. DOI: **http://dx.doi.org/10.1145/2207676.2208350**

4. Nicole David. 2012. New Frontiers in the Neuroscience of the Sense of Agency. *Frontiers in Human Neuroscience* 6 (2012), 161. DOI: **http://dx.doi.org/10.3389/fnhum.2012.00161**

5. Beth Arburn Davis. 2004. *Development and Validation of a Scale of Perceived Control Across Multiple Domains*. Ph.D. Dissertation. Philadelphia College of Osteopathic Medicine.

6. John A. Dewey and Günther Knoblich. 2014. Do Implicit and Explicit Measures of the Sense of Agency Measure the Same Thing? *PLoS One* 9, 10 (Oct. 2014), 1–9. DOI: **http://dx.doi.org/10.1371/journal.pone.0110118**

7. Mia Y. Dong, Kristian Sandberg, Bo M. Bibby, Michael N. Pedersen, and Morten Overgaard. 2015. The Development of a Sense of Control Scale. *Frontiers in Psychology* 6 (2015), 1733. DOI: **http://dx.doi.org/10.3389/fpsyg.2015.01733**

8. Patrick Haggard. 2005. Conscious Intention and Motor Cognition. *Trends in Cognitive Sciences* 9, 6 (2005), 290 – 295. DOI: **http://dx.doi.org/10.1016/j.tics.2005.04.012**

9. Sandra G. Hart. 2006. NASA-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50 (2006), 904–908. DOI: **http://dx.doi.org/10.1177/154193120605000909**

10. Pamela J. Hinds. 1998. *User Control and Its Many Facets: A Study of Perceived Control in Human-Computer Interaction*. Technical Report. Hewlett Packard Laboratories.

11. Daniel Lakens. 2013. Calculating and Reporting Effect Sizes to Facilitate Cumulative Science: A Practical Primer for T-Tests and Anovas. *Frontiers in Psychology* 4 (2013), 863. DOI: **http://dx.doi.org/10.3389/fpsyg.2013.00863**

12. Hannah Limerick, David Coyle, and James W. Moore. 2014. The Experience of Agency in Human-Computer Interactions: A Review. *Frontiers in Human Neuroscience* 8 (2014), 643. DOI: **http://dx.doi.org/10.3389/fnhum.2014.00643**

13. Hannah Limerick, James W. Moore, and David Coyle. 2015. Empirical Evidence for a Diminished Sense of Agency in Speech Interfaces. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, Seoul, Republic of Korea, 3967–3970. DOI: **http://dx.doi.org/10.1145/2702123.2702379**

14. Christopher A. Miller and Raja Parasuraman. 2007. Designing for Flexible Interaction Between Humans and Automation: Delegation Interfaces for Supervisory Control. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 49, 1 (2007), 57–75. DOI: **http://dx.doi.org/10.1518/001872007779598037**

15. Fabrizio Pece, James Tompkin, Hanspeter Pfister, Jan Kautz, and Christian Theobalt. 2014. Device Effect on Panoramic Video+Context Tasks. In *Proceedings of the 11th European Conference on Visual Media Production (CVMP '14)*. ACM, London, United Kingdom, 14:1–14:9. DOI: **http://dx.doi.org/10.1145/2668904.2668943**

16. Jeremy I. Robson and Jonathan M. Crellin. 1989. The Role of User's Perceived Control in Interface Design, Employing Verbal Protocol Analysis. *Applied Ergonomics* 20, 4 (1989), 246–251.

17. Ben Shneiderman, Catherine Plaisant, Maxine Cohen, and Steven Jacobs. 2009. *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (5th ed.). Addison-Wesley Publishing Company, USA.

18. Uri Simonsohn, Leif D. Nelson, and Joseph P. Simmons. 2014. P-Curve: A Key to the File-Drawer. *Journal of Experimental Psychology: General* 143, 2 (2014), 534–547. DOI: **http://dx.doi.org/10.1037/a0033242**

19. Matthis Synofzik, Gottfried Vosgerau, and Albert Newen. 2008. Beyond the Comparator Model: A Multifactorial Two-Step Account of Agency. *Consciousness and Cognition* 17, 1 (2008), 219–239. DOI: **http://dx.doi.org/10.1016/j.concog.2007.03.010**

20. Wen Wen, Atsushi Yamashita, and Hajime Asama. 2015. The Sense of Agency During Continuous Action: Performance Is More Important Than Action-Feedback Association. *PLoS One* 10, 4 (2015), e0125226. DOI: **http://dx.doi.org/10.1371/journal.pone.0125226**

21. Michael J. Wenger. 1991. On the Rhetorical Contract in Human—Computer Interaction. *Computers in Human Behavior* 7, 4 (1991), 245 – 262. DOI: **http://dx.doi.org/10.1016/0747-5632(91)90013-Q** Locus of Control questionnaire in Appendix D.

22. Dingyun Zhu, Tom Gedeon, and Ken Taylor. 2011. “Moving to the Centre”: A Gaze-Driven Remote Camera Control for Teleoperation. *Interacting with Computers* 23, 1 (2011), 85–95. DOI: **http://dx.doi.org/10.1016/j.intcom.2010.10.003**