Exploring Cognitive Load of Single and Mixed Mental Models Gesture Sets for UAV Navigation

Martin Hitz, Ekaterina Königstorfer, Ekaterina Peshkova

Exploring Cognitive Load of *Single* and *Mixed* Mental Models Gesture Sets for UAV Navigation

Martin Hitz
Institut für Informatik-Systeme
Alpen-Adria-Universität Klagenfurt, Austria
martin.hitz@aau.at

Ekaterina Königstorfer
Institut für Informatik-Systeme
Alpen-Adria-Universität Klagenfurt, Austria
ekzhuravleva@edu.aau.at

Ekaterina Peshkova
Institut für Informatik-Systeme
Alpen-Adria-Universität Klagenfurt, Austria
ekaterina.peshkova@aau.at

ABSTRACT

We compare four gesture sets for controlling a UAV in terms of cognitive load, intuitiveness, easiness, learnability, and memorability, by means of users' subjective feedback. Additionally, we evaluate the level of cognitive load associated with each gesture set under study using dual-task performance measures (errors and response time) and time perception. Our participants used all four gesture sets under study in a Wizard of Oz based simulated environment. Results confirm our hypothesis that *mixed mental model* gesture sets perform worse than *single mental model* gesture sets in terms of all the considered attributes. However, we did *not* find a significant difference in cognitive load between the three classes of mental models identified in our previous work.

iHDI '19 – International workshop on Human-Drone Interaction, CHI '19 Extended Abstracts, May 5, 2019, Glasgow, Scotland, UK, http://hdi.famnit.upr.si

KEYWORDS

Mental model; gesture vocabulary; user study; UAV; navigation; cognitive load; time perception; Wizard of Oz; interaction vocabulary; memorability; learnability; intuitiveness; coherence

INTRODUCTION

As sensing devices for HCI such as Kinect [9][22] and Leap Controller [12][23][26] have become affordable, a higher interest in the design of more natural and intuitive HCI has arisen, especially in Human-UAV Interaction [5][7][16][17][18][19][20][21]. We interact with machines using a wide spectrum of natural input modalities: gestures, speech, facial expressions, and gaze direction.

One of the key questions addressed in recent interaction studies is the *design of interaction vocabularies*. A typical way to design a vocabulary for controlling a device is to conduct an elicitation study to collect user suggestions and then to follow the majority principle, taking into account the most frequently suggested items to define the final interaction vocabulary. However, we consider this method insufficient to approach an "optimum" interaction vocabulary.

While several authors suggest to achieve interaction intuitiveness using different metaphors that evoke certain mental models of the system to interact with [4][5][14][15], Peshkova et al. have previously advocated the importance to **restrict commands** (vocabulary items) **to those associated with a *single* mental model** when aiming at intuitive interaction [18][20] and have grouped collected examples of such models into three classes – *instrumented*, *imitative*, and *intelligent*. The imitative class suggests that a device can imitate its operator's movements. In the instrumented class, an operator interacts via an imaginary link, e.g., an invisible joystick. In the intelligent class, a UAV is associated with an intelligent living being that can understand and follow more abstract commands.

Single **mental model** vs. ***mixed*** **mental model gesture vocabularies**: If *all* gestures within a gesture set (vocabulary) are based on a *single* underlying mental model, we call the set a *single mental model vocabulary*, otherwise *a mixed mental model vocabulary*.

The key difference between the three classes is the expectations they raise and the need for initial instruction. According to the authors' hypothesis, the intelligent class has the lowest cognitive load because a user controls a system akin to everyday interactions, and thus no additional advice is necessary. For the other classes, a user needs a hint defining the interaction characteristics (e.g., "your hand represents a UAV") – and must remember them, so the cognitive load is slightly higher, while the instrumented class requires some knowledge about the imaginary link that is used to control a UAV. Thus, in the latter case, cognitive load should be the highest. In our study, we investigate this hypothesis. For this purpose, we selected one gesture set from each class of mental models among user-defined gesture sets from a previous exploratory study [20] and decided to use the following measures to test the hypothesis: *dual-task performance*, *participants' subjective evaluation*, and *time perception*.

Time perception: A relatively new measure in HCI, but according to pilot studies, promising to represent a reliable indicator of cognitive load [2, 3, 8]: It is based on the observation that when a person focuses on a task and is actively engaged in it, the time seems to pass faster than usual, while when being occupied with something easy (and perhaps even a bit boring), the time seems to pass slower.

The second hypothesis put forward by Peshkova et al. is that a single mental model interaction vocabulary is in overall "better" compared to a mixed mental models interaction vocabulary. Therefore, we evaluated the two types of interaction vocabularies in terms of their respective intuitiveness, easiness, memorability, and learnability. We assessed these attributes through questionnaires. To create a mixed mental model gesture set, we intentionally mixed gestures from different mental models.

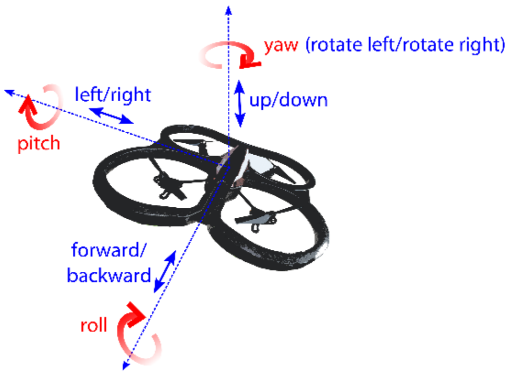

Figure 1. Moving directions, yaw, pitch, and roll axes

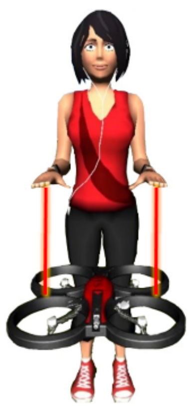

Figure 2. Neutral position for Puppeteer

GESTURE SETS

Peshkova et al. [19] investigated spontaneous gestures that non-experienced users invent to steer a UAV using basic commands (Figure 1). In a first user study, they interviewed novice users to gather their suggestions for relevant gestures for UAV navigation. In a second study, they observed spontaneous behavior of another group of novice users who were controlling the flight of a real UAV using their own gestures. As an outcome, the authors came up with a collection of gesture sets, some of which are employed in this study.

Later, Peshkova et al. [21] analyzed commonalities of the obtained gesture sets. As a result, three classes of mental models were identified: *imitative*, *instrumented*, and *intelligent* (see Introduction). For our study, we selected the *Full Body* mental model as a representative of the *imitative* class of mental models: A UAV imitates its operator's full-body movements – if you step forward, the UAV flies forward etc. The *instrumented* model class is represented by the *Puppeteer* mental model: The user carries an imagined vehicle right ahead of her/him, "linked" with the user's hands via two virtual strings, the real vehicle copies the actions of the imagined one (Figure 2). In the *intelligent* class, a user interacts with a UAV supposing that it is intelligent enough to interpret the user's "high-level" gestures. Following this idea, we asked our participants to invent their own "intelligent" gestures for basic navigation commands (in the following called *MyG*, short *for "My Gestures"*). The participants had complete freedom to use any relevant gestures under the condition that a human user controlling the UAV could interpret the invented gestures.

Figure 3 shows the three predefined gesture sets. The last row of Figure 3 presents the gestures from the *Mixed* gesture set. This set consists of gestures from diverse mental models: *Puppeteer* (*up* and *down*); *Full Body* (*forward* and *backward*); *Indication* (rotation commands); and *Airplane* (*left* and *right*: based on the "airplane" mental model). Thus, *Mixed* represents a mixed mental models gesture set as opposed to *Full Body* and *Puppeteer* which are associated with a single model each.

In our study, we investigated how the user's cognitive load depends on the employed gesture sets. We checked whether we could find a difference (1) between the three classes of mental models and (2) between *single mental model* gesture sets and *mixed mental models* gesture sets.

Based on the discussion provided earlier (see Introduction), we hypothesize that the lowest cognitive load is associated with intelligent mental models (*MyG*) and the highest with the instrumented mental models (*Puppeteer*). We expect gesture sets with gestures from *imitative* mental models (*Full Body*) to impose cognitive load higher than *intelligent* and lower than *instrumented* (H1). Our second hypothesis (H2) is that people experience higher cognitive load and lower intuitiveness, memorability, learnability, and easiness using *mixed mental models* gestures sets (*Mixed*) compared to *single mental model* gesture sets (*Full Body* and *Puppeteer*).

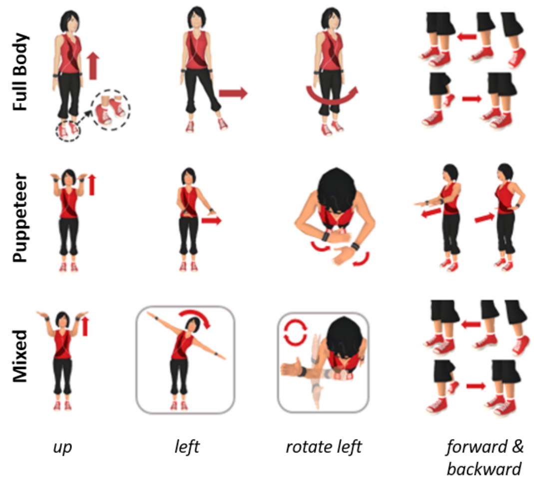

Figure 3. The three gesture sets investigated

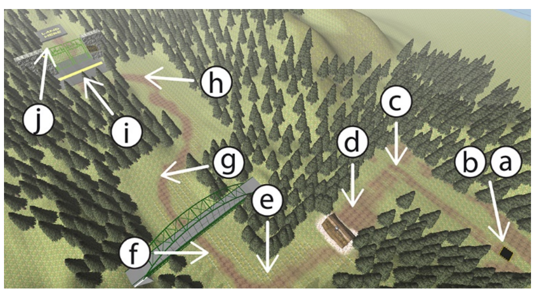

Figure 4 [20]. **Overview of Route 1**

USER STUDY

We simulated a UAV's flight using a 3D computer simulation that consists of four pre-defined flight routes of equal difficulty [20]. The 22 participants' (aged between 19 and 34 years; 6 female) task was to control the vehicle on these routes using different gesture sets. To fly along each route, the participants had to use the same ten navigation commands, but in changing order.

Figure 4 offers an outline of the first route. There are eight checkpoints between the start (a) the end point (j). Providing the appropriate commands, the user crosses all checkpoints and reaches the destination (j) where the vehicle is supposed to land.

In order to measure the participants' baseline time perception, we recorded the time the participants felt to constitute one minute.

After having watched a short video of one of the four routes, each participant performed the navigation task four times, first with set *MyG* and then once with each pre-defined gesture set (counterbalanced to prevent problems with sequence effects [11]): *Full Body*, *Puppeteer*, and *Mixed*.

We collected users' time perception and their subjective evaluation of cognitive load experienced when using different gesture sets. The participants also reflected their subjective evaluation of the used gesture sets in a questionnaire before proceeding with a new gesture set. They answered the questions in regard to cognitive load (7-point scale) and time perception (how long it took to finish the route in their opinion). During the entire experiment, the experimenter took notes about think aloud data. When the participants completed the tasks, we asked them to evaluate the four gesture sets in terms of their intuitiveness, easiness, and memorability. In the final questionnaire the participants also gave a description of the gestures and selected their favorite/least favorite gesture set(s). The participants were also asked to explain their choice.

Before starting the navigation task with each of the pre-defined sets, the experimenter showed all the gestures one by one (*Mixed*) and also explained the underlying idea of the single mental model gesture sets (*Full Body* and *Puppeteer*). Moreover, the participants received an instruction sheet that showed all gestures (see **Figure 3**). The participants could take as long as they required to study the gestures before proceeding to the task execution. As we observed, participants spent no time studying the instruction sheets and started steering the UAV right after the explanation (a couple of participants took a few seconds to review gestures from set *Mixed*). For the duration of the task, the participant could not look into the list of gestures.

During each navigation task, the experimenter asked participants five simple math questions ("3+2=?", "2x4=?", etc.), wrote down the participants' answers, recorded time delays (when the response time was more than 5 seconds) and wrong answers, and took notes about think-aloud data. These math questions represented the second task that our participants had to perform simultaneously with the main navigation task. The information regarding time delays and wrong answers is intended to reflect the participants' cognitive load.

A video recording explaining the study can be found on YouTube [29].

Gesture set	**Min.**	**Med.**	**Mean**	**Max.**	**S.D.**
MyG	-68.5	43.48	62.24	255.88	83.31
Full Body	-99.31	29.86	24.06	166.59	58.79
Puppeteer	-62.66	21.33	26.73	128.27	56.15
Mixed	-64.19	22.13	24.74	146.29	58.34

Table 1. Descriptive statistics of the error of time perception: Estimated Time – Actual Time

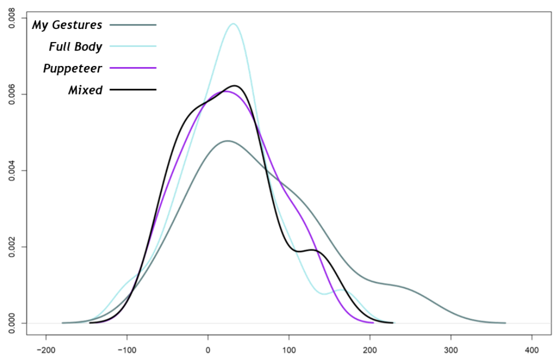

Figure 5. Densities of time deviations

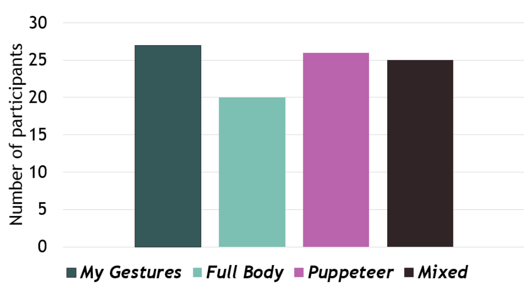

Figure 6. Delays and wrong answers

RESULTS

Cognitive load

To evaluate the level of cognitive load, we used time perception, dual-task performance, and participants' subjective evaluation [1][2][4][8][10][24][26][28]. We also assessed intuitiveness, easiness, memorability, and learnability of the considered gesture sets through questionnaires.

Time perception: After performing the navigation task with each gesture set, our participants guessed the time spent to complete the task. Block & Gellersen explored the influence of cognitive load on the perception of time [2]. It has been found that an increase of cognitive load leads to a decrease in time perception [1]. Hart [8] and Zakay & Shub [28] discovered that participants usually underestimated time intervals when the task load was higher [1]. The descriptive statistics for the error in time perception for each gesture set is presented in Table 1. From Figure 5 we can see that some participants notably overestimated the time spent with *MyG*. We conducted Friedman's test [13] and found that the main effect of gesture set tended to be significant: $\chi^2(3)$ = 7.25, p = 0.06. Overall, we observed an overestimation of time (Table 1). The participants perceived the time spent with *MyG* longer than with other gesture sets. The number of participants who underestimated the time was: 5 (*MyG*), 7 (*Full Body*), 8 (*Puppeteer* and *Mixed*). That supports (though not significantly) our hypothesis that the cognitive load associated with *intelligent* gesture set (*MyG*) was the lowest, with *imitative* (*Full Body*) slightly higher, and the highest with *instrumented* (*Puppeteer*).

Dual-Task: We counted how many delays and wrong answers to math questions the participants made while steering the UAV. Figure 6 shows the obtained results. We did not find significant differences between the four gesture sets (Friedman's test: $\chi^2(3)$ = 1.46, p = 0.69).

Subjective Evaluation: Figure 7 shows the results of the subjective evaluation of cognitive load experienced using 7-point scale (1 – very low, 7 – very high). The most frequent evaluation (mode) for *MyG* was 3, perhaps because of the fact that it was always the first set. *Full Body* and *Puppeteer* were most frequently evaluated as 1 and 2, respectively. The most frequent evaluation for the *Mixed* gesture set was 4, implying that this set was perceived the most complicated. However, the difference between the four sets was not significant (Friedman's test: $\chi^2(3)$ = 4.38, p = 0.22).

Learnability

At the end of the experiment, we asked the participants to write down a description of each gesture set for the next participant who would not see the actual gestures, but control the flight using the written description. They should either describe each gesture individually *or* describe the idea behind each set if they consider it sufficient to complete the navigation task. As shown in Figure 8, 5 and 11 participants decided that it is enough to give a hint (the "main idea") to describe *Full Body* and *Puppeteer* set, respectively. The majority of participants gave a full description for *MyG*, perhaps because they had not enough time, or they did not recognize an idea behind their own gestures. As expected, all participants gave a full description for *Mixed*.

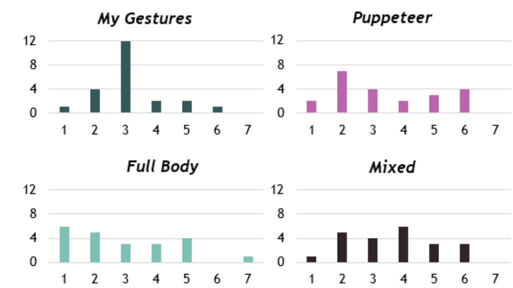

Figure 7. Subjective evaluation of cognitive load

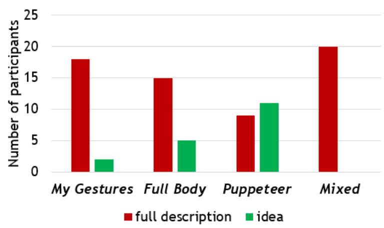

Figure 8. Description of gesture sets

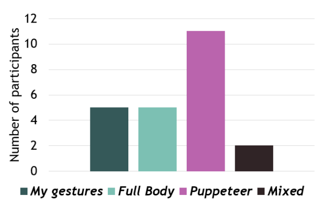

Figure 9. Favorite gesture set

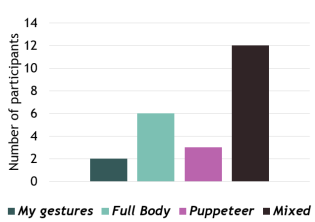

Figure 10. Least-liked gesture set

Priorities

After completion of all tasks, we asked the participants to choose their favorite and least-liked gesture sets (multiple answers allowed). Participants also ordered the four gesture sets based on their intuitiveness, easiness, and memorability. We analyzed the differences between the subjective evaluations with Friedman's test. A pairwise Wilcoxon test with Bonferroni correction was used for the post-hoc analysis. *Puppeteer* was favorite of most of the participants while *Mixed* was disliked most (Figure 9, Figure 10). 5 participants mentioned that they got the best impression from their own gestures and *Full Body*. 12 participants scored *Mixed* as least-liked. A significantly greater number of participants found the *single mental model* gesture sets (*Full Body* and *Puppeteer*) *more intuitive* compared to the *mixed mental models* gesture set (*Mixed*): $\chi^2(3) = 12.90$, $p = 0.005$; post-hoc for *Mixed* with *Full Body* and *Mixed* with *Puppeteer:* $p = 0.007$, $p = 0.058$, respectively. *Mixed* was evaluated significantly *more complicated* than the other gesture sets: $\chi^2(3) = 17.12$, $p = 0.0007$; post-hoc (*MyG*): $p = 0.008$; post-hoc (*Full Body*): $p = 0.0002$; post-hoc (*Puppeteer*): $p = 0.0042$. *Mixed* was also evaluated significantly *less memorable* than the other gesture sets: $\chi^2(3) = 17.08$, p 0.00068; post-hoc (*MyG*): $p = 0.02$; post-hoc (*Full Body*): $p = 0.001$; post-hoc (*Puppeteer*): $p = 0.003$.

DISCUSSION

We did not find significant differences between gesture sets in terms of cognitive load. However, we did observe some notable differences between the four gesture sets. Specifically, based on our time perception measures, we noticed that set *MyG* was associated with the lowest cognitive load indicator among the four sets under study. *MyG* represents the *intelligent* class of mental models: consisting of gestures borrowed from human-to-human interaction. Though the participants had complete freedom to suggest gestures, we did not find much variety among their behavior. Basically, their gestures could be described via a single sentence: "Use your hand to indicate the direction to fly or rotate." Thus, the participants tended to follow a single idea and their gestures actually adhere to a single mental model – which constitutes another interesting finding.

Overall, *Mixed* received the worst evaluation compared to the other three sets, thus supporting hypothesis H2 and previous research [20]. As a result, this set was selected by the majority of participants as the least-liked one. Considering that we intentionally selected gestures from different mental models for this gesture set, the obtained result is not really surprising, but it does stress the importance of adhering to a single mental model when designing a gesture-based vocabulary.

Though the obtained results do not formally support our first hypothesis (H1, cognitive load grows from *intelligent* over *imitative* to *instrumented* gesture sets: *MyG* < *Full Body* < *Puppeteer*), we did observe some tendency in favor of this hypothesis. Thus, it seems promising to us to further investigate cognitive load associated with different classes of mental models using more precise measures, such as pupil dilation, and to consider a couple of representatives from each class of mental models for a more comprehensive comparison. However, as shown by E et al., due to cultural differences, in any case, we cannot expect a single "one size fits all" solution [7].

REFERENCES

[1] Hee-Kyung Ahn and Maggie Wenjing Liu and Dilip Soman. 2006. Memory for Time: A Cognitive Model of Retrospective Duration and Sequence Judgments. In SSRN Electronic Journal. http://dx.doi.org/10.2139/ssrn.897933

[2] Daniel Baldauf, Esther Burgard and Marc Wittmann. 2009. Time perception as a workload measure in simulated car driving. In *Applied Ergonomics*. 40(5), 929–935. http://dx.doi.org/10.1016/j.apergo.2009.01.004

[3] Florian Block and Hans Gellersen. 2010. The impact of cognitive load on the perception of time. In *Proc. of the 6th Nordic Conference on Human-Computer Interaction Extending Boundaries* (NordiCHI '10). Available at: http://dx.doi.org/10.1145/1868914.1868985

[4] John M. Carroll and Judith Reitman Olson. 1988. Mental models in human-computer interaction. In *Handbook of Human-Computer Interaction* (1st. ed.), Martin G. Helander (ed.). The Netherlands: Elsevier Science B. V., 45-65. https://doi.org/10.1016/b978-044481862-1/50069-8

[5] Jessica R. Cauchard, Jane L. E, Kevin Y. Zhai, & James A. Landay. 2015. Drone & me: an exploration into natural human-drone interaction. In *Proc. 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15,* 361–365. New York, NY, USA: ACM Press. https://doi.org/10.1145/2750858.2805823

[6] Krista E. DeLeeuw and Richard E. Mayer. 2008. A Comparison of Three Measures of Cognitive Load: Evidence for Separable Measures of Intrinsic, Extraneous, and Germane Load. In *Journal of Educational Psychology*. 100, no. 1: 223–234. doi:10.1037/0022-0663.100.1.223.

[7] Jane L. E, Ilene L. E, James A. Landay, and Jessica R. Cauchard. 2017. In *Proc. 2017 CHI Conf. on Human Factors in Computing Systems, Denver, CO, USA — May 06 - 11, 2017*, 6794-6799, ACM Press New York, NY.

[8] Sandra G. Hart. 1978. Time estimation as a secondary task to measure workload: summary of research. In: *Proceedings of the 14th Annual Conference on Manual Control (NASA CD-2060)*. NASA, Washington, DC, 82–90

[9] Kohki Ikeuchi, Tomoaki Otsuka, Akihito Yoshii, Mizuki Sakamoto, and Tatsuo Nakajima. 2014. KinecDrone: enhancing somatic sensation to fly in the sky with Kinect and AR.Drone. In *Proceedings of the 5th Augmented Human International Conference* (AH '14). https://doi.org/10.1145/2582051.2582104

[10] Azizuddin Khan, Narendra K. Sharma, and Shikha Dixit. 2006. Effect of cognitive load and paradigm on time perception. In *Journal of Indian Academy of Applied Psychology*. 32 (1), 37-42.

[11] Beob Kim and Hans Stein. 2009. A spreadsheet program for making a balanced Latin Square design. In *Revista Colombiana de Ciencias Pecuarias*. 22. 591-596.

[12] Daniel Liebeskind. 2013. Leap Motion Node Drone Flight in 2013. Video. (19 November, 2013.). Retrieved January 26, 2017 from https://www.youtube.com/watch?v=hfq2SisPvCU

[13] Evie McCrum-Gardner. 2008. Which is the correct statistical test to use? In *British Journal of Oral and Maxillofacial Surgery*. 46(1), 38–41. http://dx.doi.org/10.1016/j.bjoms.2007.09.002

[14] Dennis C. Neale and John M. Carroll. 1997. The role of metaphors in user interface design. In *Handbook of Human-Computer Interaction* (2nd. ed.), Martin G. Helander, Thomas K. Landauer, and Prasad V. Prabhu (eds.). The Netherlands: Elsevier Science, B. V., 441-462. https://doi.org/10.1016/b978-044481862-1/50086-8

[15] Norman, D. A. (1987). Some observations on mental models. In *Human-Computer Interaction* (241-244). Morgan Kaufmann Publishers Inc. San Francisco, CA, USA.

[16] Mohammad Obaid, Felix Kistler, Gabrielė Kasparavičiūtė, Asım Evren Yantaç, and Morten Fjeld. 2016. How Would You Gesture Navigate a Drone? A User-Centered Approach to Control a Drone. In *Proc. 20th International Academic Mindtrek Conference, Tampere, Finland — October 17 - 18, 2016*, 113-121. ACM Press New York, NY.

[17] Ekaterina Peshkova. 2017. *Approaching Naturalness and Intuitiveness of Input Vocabularies for UAV Navigation*. Ph.D Dissertation. Alpen-Adria-Universität Klagenfurt, Fakultät für Technische Wissenschaften.

[18] Ekaterina Peshkova and Martin Hitz. 2017. Coherence evaluation of input vocabularies to enhance usability and user experience. In *Proceedings of the 9th ACM SIGCHI Symposium on Engineering Interactive Computing Systems* (EICS '17). http://dx.doi.org/10.1145/3102113.3102118

[19] Ekaterina Peshkova, Martin Hitz, and David Ahlström. 2016. Exploring user-defined gestures and voice commands to control an unmanned aerial vehicle. In *Proceedings of the 8th International Conference on Intelligent Technologies for Interactive Entertainment* (INTETAIN '16), 47-62. http://dx.doi.org/10.1007/978-3-319-49616-0_5

[20] Ekaterina Peshkova, Martin Hitz, David Ahlström, Rainer W. Alexandrowicz, and Alexander Kopper. 2017. Exploring intuitiveness of metaphor-based gestures for UAV navigation. In *Proceedings of the 26th IEEE International Symposium on Robot and Human Interactive Communication* (RO-MAN '17). http://dx.doi.org/10.1109/roman.2017.8172298

[21] Ekaterina Peshkova, Martin Hitz, and Bonifaz Kaufmann. 2017. Natural interaction techniques for an unmanned aerial vehicle system. *IEEE Pervasive Computing* 16, 1: 34-42. https://doi.org/10.1109/mprv.2017.3

[22] Kevin Pfeil, Seng Lee Koh, Joseph LaViola. 2013. Exploring 3d gesture metaphors for interaction with unmanned aerial vehicles. In *Proceedings of the International Conference on Intelligent User Interfaces* (IUI '13), 257-266. https://doi.org/10.1145/2449396.2449429

[23] Ayanava Sarkar, Ketul Arvindbhai Patel, R. K. Ganesh Ram, and Geet Krishna Capoor. 2016. Gesture control of drone using a motion controller. In *Proceedings of the 2016 International Conference on Industrial Informatics and Computer Systems* (CIICS '16). https://doi.org/10.1109/iccsii.2016.7462401

[24] Annett Schmeck, Maria Opfermann, Tamara van Gog, Fred Paas, Detlev Leutner. 2013. Measuring cognitive load with subjective rating scales during problem solving: differences between immediate and delayed ratings. In *Instructional Science*. 43(1), 93–114. http://dx.doi.org/10.1007/s11251-014-9328-3

[25] Alexander Skulmowski and Günter Daniel Rey. 2017. Measuring Cognitive Load in Embodied Learning Settings. In *Frontiers in Psychology*, 8. http://dx.doi.org/10.3389/fpsyg.2017.01191

[26] Ramon A. Suarez Fernandez, Jose Luis Sanchez-Lopez, Carlos Sampedro, Hriday Bavle, Martin Molina, and Pascual Campoy. 2016. Natural user interfaces for human-drone multi-modal interaction. In *Proc. of the 2016 International Conference on Unmanned Aircraft Systems* (ICUAS '16), 1013-1022. https://doi.org/10.1109/icuas.2016.7502665

[27] Geert Wernaart. 2013. Cognitive load measurement: Different instruments for different types of load? Master Thesis. Utrecht: Faculty of Social and Behavioral Sciences.

[28] Dan Zakay and Joseph Shub. 1998. Concurrent duration production as a workload measure. In *Ergonomics*. 41(8), 1115-1128. http://dx.doi.org/10.1080/001401398186423

[29] [Interactive Systems]. (2018, July 9). User Study. Retrieved from https://www.youtube.com/watch?v=W4a8GnULDFo