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

Ekaterina Königstorfer
Alpen-Adria-Universität Klagenfurt
Klagenfurt am Wörtersee, Austria
{ekzhuravleva}@edu.aau.at

ABSTRACT

We conducted a user study to compare four gesture sets in terms of cognitive load, intuitiveness, easiness, learnability, and memorability. We assessed these attributes through the users' subjective feedback reflected in questionnaires. Additionally, to evaluate the level of cognitive load associated with each gesture set under study, we used dual-task performance measures (errors and response time) and a relatively new for Human-Computer Interaction measure – time perception. In our study, participants were controlling the flight of a UAV in a simulated environment using all four gesture sets. We adapted the Wizard of Oz approach in the study design: the flight was actually controlled by a human operator. The collected data confirmed our hypothesis that *mixed mental model* gesture sets are worse than *single mental model* gesture sets in terms of all the considered attributes. However, we did not find a significant difference in terms of cognitive load between the three classes of mental models (*intelligent*, *imitative*, and *instrumented*) only a tendency towards our hypothesis.

Author Keywords

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

ACM Classification Keywords

H.5.2. User Interfaces: Evaluation/Methodology, Interaction styles, User-centered design.

INTRODUCTION

Nowadays, we observe how recently released sensing devices for Human-Computer Interaction (HCI) such as Kinect [7, 19] and Leap Controller [10, 20, 23] are gaining more and more popularity among users. As these devices become affordable for a wider range of researchers and practitioners, a higher interest in the design of more natural and intuitive HCI (including Human-Robot Interaction and Human-UAV Interaction) arises. Nowadays, we can interact with machines not only via standard input devices such as a keyboard and a mouse, but also using a wide spectrum of natural input modalities: gestures, speech, facial expressions, and gaze direction. We observe a growing amount of research works in this direction.

One of the key questions addressed in the recent research works is the design on an interaction vocabulary. A typical way to design the vocabulary is to conduct an elicitation study to collect users suggestions (e.g., user-defined gestures and voice commands) and then to follow the majority principle (that suggests that the most frequently observed gestures/voice commands are selected for the final vocabulary) to define the final interaction vocabulary. But is this method indeed a good way to achieve a 'better' interaction vocabulary?

Previously, Peshkova et al. [15, 17] have shown the importance of considering the adherence to a single mental model as an important criterion when aiming at intuitive interaction. The authors suggested a new classification scheme that is based on the concept of mental model. Mental models are formed by previously gained knowledge and experiences. Thus, each mental model is associated with certain knowledge. In HCI, users employ their mental models of a certain system to correctly interact with this system, predict its behavior, and correct possible errors. Overall, mental models allow their owners to choose a proper way to behave [14].

Many of the works [4, 12, 13] try to achieve interaction intuitiveness using different metaphors that evoke certain mental models. The authors clustered supportive examples of these models into three categories – *imitative*, *instrumented*, and *intelligent*.

The *imitative* class implies that a device can imitate operator's movements. This interaction can be seen as a direct mapping of the operator's movements to the vehicle motion, e.g., the 'hand' mental model, in which operator's hand represents a UAV, thus, the operated UAV simply copies movements of the operator's hand.

The *instrumented* class suggests that an operator controls a vehicle through an imaginary intermediate link, which can be an imaginary physical object, such as a joystick; an imaginary link, such as an invisible string that lets you

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

manipulate the vehicle as if it were a marionette; or a super force that lets you move the vehicle without touching it, such as repelling or attracting a vehicle with an open palm.

The key feature of the *intelligent* class, as its name implies, is that a UAV is treated as an intelligent creature. This explains the fact that, in many cases, this class is deemed most likely to resemble natural interaction. For example, when a person is instructed where to go, people tend to describe a place verbally and redundantly point in a direction with their index finger.

The key difference between the presented classes is the expectations they raise and the need for initial instruction. The authors hypothesize that the intelligent class has the lowest cognitive load because an operator navigates a system following the natural 'flair' without a need for prior instructions. For the imitative and instrumented classes, a hint specifying the type of interaction is needed, so the cognitive load is higher. In addition, the instrumented class requires certain knowledge and experience from an operator. Thus, 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 the previous exploratory study [17].

After a careful literature overview [5, 21, 22, 24], we decided to use the following standard measures to test this hypothesis: dual-task performance and participants' subjective evaluation. We also decided to apply an extra measure – time perception. It is a relatively new measure in HCI, but based on the pilot works, it seems to be a reliable indicator of cognitive load [2, 3, 8]. It is believed when a person focuses on something and is actively engaged in some task, the time seems to pass faster than it actually does. Whereas when a person does something easy and perhaps even a bit boring, the time seems to pass slower.

The second hypothesis that the authors rise is that a *single mental model* interaction vocabulary is in overall 'better' compared to a *mixed mental models* interaction vocabulary. Apart from cognitive load, in the current study, we compared these 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.

In our study, participants controlled the flight of a UAV in a simulated environment. Each participant completed the navigation task once with each gesture set under study (see Gesture Sets). We employed the Wizard of Oz technique in order to get rid of potential errors of a gesture recognizer. Our Wizard was uncovered: the participants knew that a human operator was actually controlling the simulated UAV based on their gestures. The participants were aware that the operator can only see them and not hear.

We start by describing the four gesture sets investigated in this work. Then, we present the study design and show the obtained results. Afterwards, we discuss our findings and suggest for the direction for a future research. In conclusion, we summarize the outcomes of the study.

GESTURE SETS

In the previous study, Peshkova et al. [16] explored intuitive behavior of novice users for UAV navigation. The focus of their study was on basic navigation commands (as in standard off-the-shelf UAVs) to steer a UAV that include *up*, *down*, *left*, *right*, *rotate left*, *rotate right*, *forward*, and *backward* (Figure 1). As an outcome, the authors first collected users' suggestions for relevant gestures through interview sessions and then recorded user-defined gestures showed by another group of users in a Wizard of Oz experiment, in which the users were made to believe they were steering a UAV. For a deeper understanding of users' gestures, the authors analyzed the collected data in terms of mental models. In order to identify users' mental models that likely guided users in their choice of gestures, the authors detected those mental models that allow to define most of the gestures showed by each user. As a result, the authors came up with a collection of gesture sets, in which each set is associated with a single mental model. In our work, we employ gesture sets from this collection.

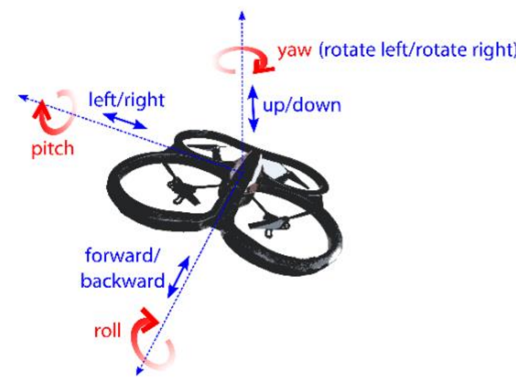

Figure 1. Moving directions, yaw, pitch, and roll axes.

Later, Peshkova et al. [18] analyzed the identified mental models in terms of their commonalities and differences. This analysis led the authors to clustering of related mental models into three classes: *imitative*, *intelligent*, and *instrumented*. In the *imitative* class, a part of the operator's body e.g., a hand, serves as a surrogate of the UAV and thus movements of this body part are directly mapped to movements of the UAV. Gestures include those where the UAV follows the motions of the head, one hand, two hands, and the upper/full body. We selected *Full Body* mental model as a representative of this class.

In the *instrumented* class, an operator gives the flight instructions using an imaginary object (e.g., a device or a tool). This class is represented through the *Puppeteer* mental model. The operator holds an imaginary UAV right in front of the body. Operator's hands are 'connected' with the UAV through two invisible 'strings', the real UAV copies the movements of the 'puppet' UAV (Figure 2).

In the *intelligent* class, an operator expects a certain level of intelligence of a UAV that enables the UAV to interpret a given command correctly. Gestures include those where an indication of the direction is given with the index finger, thumb, hand, forearm, arm, and 'come to me' or 'go away' gestures. Following this idea, we asked our participants to invent their own gestures for basic navigation commands (Figure 1). In our study, the participants had a complete freedom to use any relevant gestures under a condition that a human operator should be able to interpret the invented gestures. Throughout this paper we call the user-defined gesture set *My Gestures*. This set represents the intelligent class of mental model.

In our study, we investigate how the user's cognitive load depends on the employed gesture sets. In particular, we check whether there is a difference (1) between the three classes of mental models and (2) between *single mental model* gesture sets and *mixed mental models* gesture sets.

Our study participants used four different gesture sets to control the flight of a UAV. Figure 3 shows three of these gesture sets. The top and the second rows show the *Full Body* and *Puppeteer* gesture sets. *Full Body* is design around an idea that your full body represents a UAV: if you step forward – the UAV flies forward, if you step back – the UAV flies back, etc. *Puppeteer* is based on a 'puppeteer' mental model.

Figure 2. Neutral position for *Puppeteer*.

The last row shows the gestures in the *Mixed* gesture set, in which we intentionally mixed gestures from different mental models: *Puppeteer* (*up* and *down*); *Full Body* (*forward* and *backward*); *Indication* (*rotate left* and *rotate right*: a rotation of the forearm at the elbow indicates the direction to rotate); and *Airplane* (*left* and *right*: designed on the 'airplane' mental model). Thus, *Mixed* represents a mixed mental models gesture set as opposed to *Full Body* and *Puppeteer* that are associated with a single model each.

Based on the discussion provided earlier (see Introduction), we hypothesize that the lowest cognitive load is associated with intelligent mental models (*My Gestures*) and the highest with the instrumented mental models (*Puppeteer*). Gesture sets with gestures from imitative mental models (*Full Body*) we expect 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*).

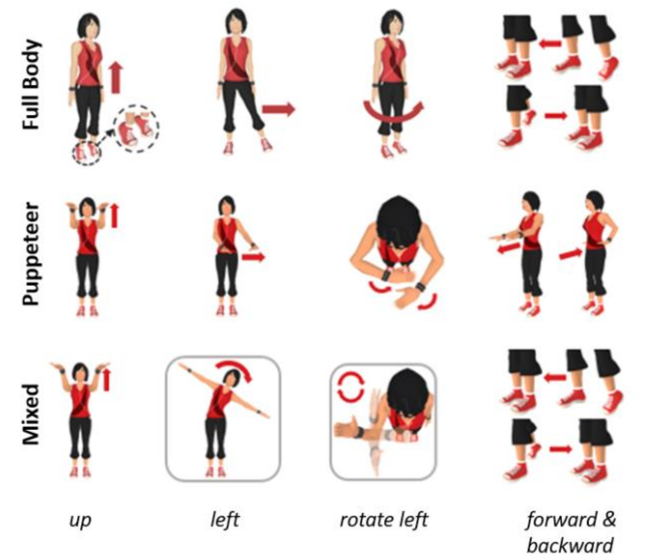

Figure 3. The three gesture sets investigated in the study.

USER STUDY

Task

To avoid damaging a real UAV and causing participants' discomfort when crashing a real UAV, we decided to simulate a UAV and its flight environment in software. We developed a 3D computer simulator that consists of four pre-defined flight routes. The user's task is to steer the UAV with gestures along these routes. Each route requires the same ten navigation commands to steer the UAV, but in different orders, as listed in Table 1.

Route 1	Route 2	Route 3	Route 4
takeoff+fw	takeoff+fw	takeoff+fw	takeoff+fw
rotate left+fw	backward+fw	right+fw	down+fw
up+fw	left+forw	rotate right+fw	rotate right+fw
rotate right+fw	rotate right+fw	down+fw	left+fw
down+fw	up+fw	left+fw	rotate left+fw
right+fw	right+fw	backward+fw	backward+fw
left+fw	rotate left+fw	rotate left+fw	up+fw
backward+fw	down+fw	up+fw	right+fw
land	land	land	land

Table 1. Command sequences in the four routes (fw=forward).

Figure 5 provides an overview of Route 1. Between the starting point (a) and the destination (j), the UAV has to pass by eight checkpoints (b-i). When the UAV approaches a checkpoint, an arrow indicating the direction to move appears next to it. This arrow serves as a hint to prompt the user to issue a new navigation command in order to avoid crashing into an obstacle. Once the UAV has crossed the checkpoint, the arrow disappears.

Figure 5 shows the user's view at all the checkpoints including the starting point and the destination (text insets with command names are not shown to the user). The following example applies to Route 1: after a successful *takeoff* command (a), the user has to issue the *forward* command (b) to move the UAV to the next checkpoint (c). At this checkpoint, the *rotate left* command is required. Then, the user issues the forward command (this and all following *forward* commands are omitted in Figure 5) to fly towards the house (d). Here the user issues the *up* command to fly over the house until the next turn (e), where the *rotate right* command is prompted. After the rotation, the UAV moves towards the upcoming bridge (f), where the user has to use the *down* command to fly the UAV below the bridge until its next turn (g), where the path makes a smooth curve. Accordingly, only a short movement to the right is necessary to get around the curve. Thereafter, the user continues with the *forward* command until the UAV approaches the next curve (h). Again, only a short sideways movement, now to the left, is required to get around the curve and face the gate (i). When the UAV faces the gate, a gatekeeper's voice demands the user to move the UAV up to the yellow line in front of the gate. At the line (once the UAV is 'identified'), the gate slowly opens, towards the UAV. Thus, the UAV has to move backward to avoid crashing into the opening doors. Once the gate is opened, the user moves the UAV to the destination platform (j) and then lands it.

A countdown timer starts when the user crosses a checkpoint. In the allotted time frame the user first has to command the UAV to fly forward to reach the next obstacle and then give a new command to pass this obstacle. Between checkpoints the user has 60 seconds when the rotate left or rotate right commands are required and 30 seconds for the other commands. Five seconds before a timeout, a warning message is displayed. If a timeout is triggered or a crash occurs, the UAV is automatically positioned directly after the current checkpoint (e.g., if the UAV crashes into the house, the UAV is positioned after it).

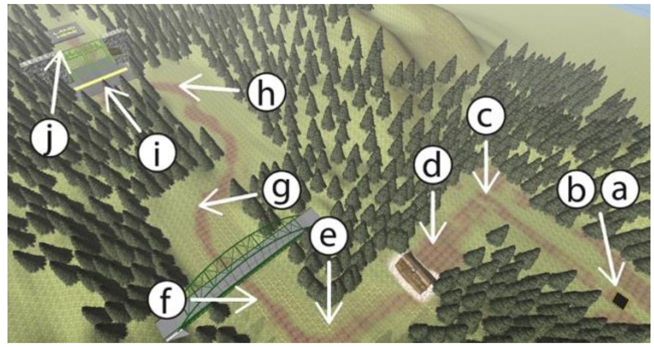

Figure 5. Route overview. Letters indicate locations of checkpoints and correspond to Figure 5.

Before proceeding with the main study, our operator practiced controlling the simulated drone using direct video streaming from the study room (two one-hour sessions). Then, we conducted a pilot study with 3 participants (3 male, from 20 to 21 years old) to estimate the study time and test our study design (the entire study procedure, technical equipment, and questionnaires). In the very first navigation task, in which the participants used their own gestures, the pilot participants were told that the operator controls the flight of a simulated UAV based on the gestures that the participants show. In the remaining navigation tasks, in which three pre-defined gesture sets were used, the participants were told that their gestures are recognized by a gesture recognizer using data from Kinect. In reality, the operator controlled the flight during all four navigation tasks.

As an outcome of the pilot study, (1) we ensured that the instructions given to participants are comprehensible and that (2) the employed equipment functions properly, (3) we

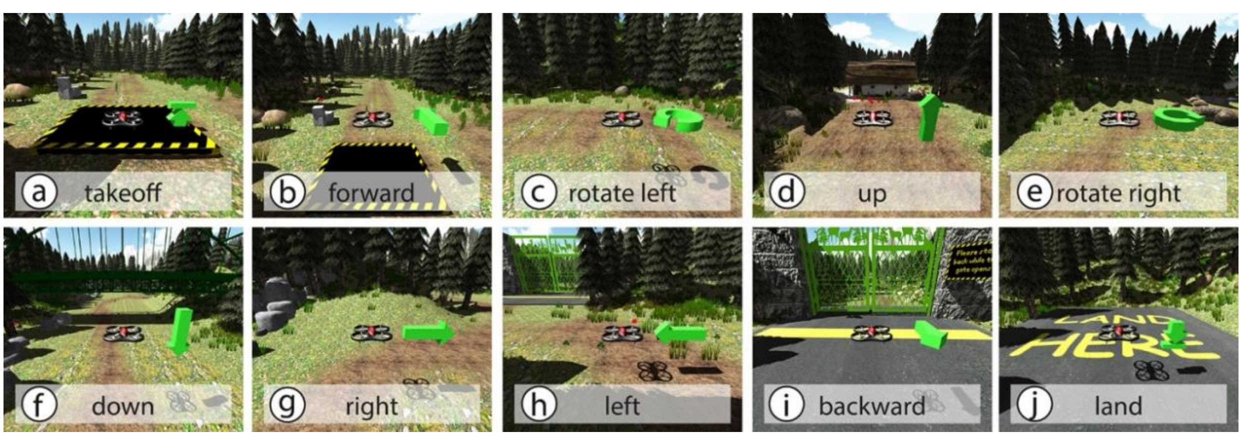

Figure 4. Vertically cropped user views: a) & b) start location, a) to j) checkpoints, j) destination (text insets invisible to the user).

re-formulated and re-arranged some questions to shorten the time required to answer them, (4) we estimated the study time (approx. 45 minutes), and (5) we decided to switch to uncovered Wizard of Oz (in which a participant is aware that the operator always controls the flight based on the participant's gestures.

During the main study, as mentioned earlier, we used the concept of uncovered 'wizard' – the participants knew that the human operator interprets their gestures and then sends the corresponding command to the flight simulator. It is worth mentioning that we had the same operator throughout the entire user study. The operator saw only participants and had no access to the video from the flight simulator.

During each navigation task, the experimenter asked participants 5 simple math questions (Table 2), 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 were the send task that our participants had to perform simultaneously with the main navigation task (a dual-task). The information regarding time delays and wrong answers is intended to reflect participants' cognitive load.

Task no. 1	Task no. 2	Task no. 3	Task no. 4
3 + 2	3 + 6	2 + 3	6 + 3
2 * 3	2 * 4	3 * 2	4 * 2
8 / 2	6 / 2	4 / 2	2 / 2
8 - 5	5 - 3	9 - 4	7 - 5
5 + 2	6 + 1	7 + 3	2 + 1

Table 2. Math questions.

Subjects

We recruited 22 participants (6 female), aged between 19 and 34 years (mean 22, S.D. 4). Nineteen participants had experience playing computer games: 9 of them played 2D and 3D games and 7 - 3D games only. Twelve participants had experience in steering remote-controlled devices (cars, boats or drones) on the yearly (10), monthly (1), and daily (1) basis. Nine participants played Kinect or Wii games every week (3), every month (3), and every year (3). Most of our participants (16) had driver's license and one of them had boating license.

Apparatus and Setup

Figure 6 shows top view of the study setup. During the task execution, the participant was standing in an upright position in front of a beamer projection, 4 meters away from the wall (projection size 1.1×0.8 meters). The video camera 1 (C1) – that is facing the participant – captures and recorders the participant's body movements and passes them to the operator's screen (number 1 in Figure 6), who sits in the adjacent room. A filled red circle and a horizontal blue dotted line on the floor mark the desired standing position respectively the minimum distance from the camera (so that the operator can observe the entire participant). When the participant shows a valid gesture, the operator sends the corresponding command to the flight simulator using a keyboard control (written using the Unity game engine, version 4.5.5). The flight simulator runs on the laptop in the operator's room (number 2 in Figure 6).

Two loudspeakers provided audio output (e.g., motor sound and crash sounds) and the other video camera 2 (C2) (facing the projection) recorded the flight simulator screen projection for later analyses.

The operator, who actually steers the simulated UAV, sits in the adjacent room (Figure 6: the room on the top). The operator observes participant's movements through the screen that receives video from camera 1 (C1). Based on the participant's gestures the operator sends one of the basic navigation commands to the simulated UAV via the keyboard control. The operator does not know what exactly is going on in the flight simulator that runs on a separate laptop (number 2 in Figure 6). The operator's keyboard and the projector (P) that displays the simulator on the projection wall are connected to this laptop. The order of routes was counterbalanced (balanced Latin square design), the operator did not know the order of routes to avoid the operator's expectation of the participants' commands.

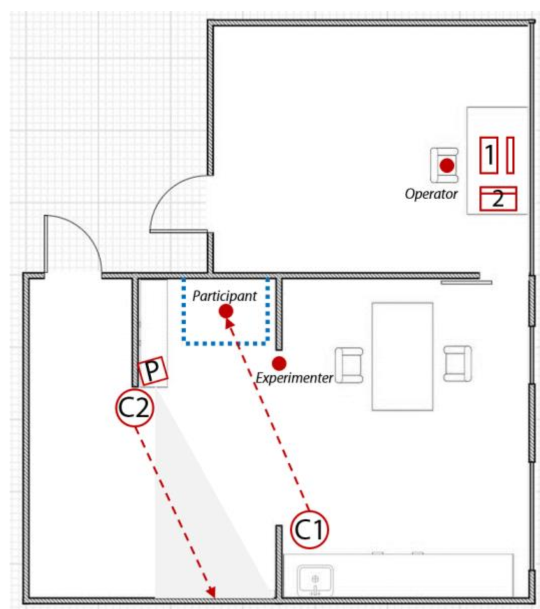

Figure 6. Study setup: top view.

Figure 7a shows an experimenter and a participant during the introduction to the study, in which the experimenter explains the study procedure, the tasks, and shows the video of one of the routes. After that, the participant steers the UAV using the four gesture sets one by one and the experimenter asks the math questions and takes notes about think-aloud data (Figure 7b). Meanwhile, the operator observes the participant movements and controls the flight correspondingly. She does not see the screen with a flight simulator and relies only on the participant's gestures (Figure 7c). To watch the video of the study follow this link [27].

Figure 7. Screenshots.

After each task, the participants completed a paper-based questionnaire, in which they evaluated the level of cognitive load they experienced during the time execution, wrote down his/her time estimation of the task completion, and evaluated their agreement with several statements using 4-point Likert scale. The participants were also encouraged to leave other relevant comments.

Study Procedure

Before the experiment, the participants read and signed the informed consent form that provided relevant study information necessary to allow the participants to make an informed and voluntary decision to participate in the experiment. They were informed about the experiment's purposes and agreed to be video recorded during their task completion for the further data analysis.

After signing the consent form, each participant received an introduction to the study procedure, the tasks, and a brief definition of cognitive load.

In order to measure participants' baseline time perception, the experimenter started the timer and asked the participants to say STOP when they think one minute is up. During this 'minute' the experimenter was asking simple math questions (similar to those used during the time execution, see Table 2). This procedure was also included into our user study to give the participants an idea on what questions they should expect during the UAV navigation. Then, the experimenter asked the participants to complete a pre-questionnaire to collect demographic data and information about their experience in related activities.

After having watched a short video (1 minute 20 seconds) that showed one of the four routes, each participant performed the navigation task four times [26]. Once with a gesture set invented by themselves (called *My Gestures*) and once with each pre-defined gesture sets: *Full Body*, *Puppeteer*, and *Mixed*.

All the participants started with *My Gestures* (they could use any gestures they find relevant for the considered navigation commands). However, they were asked to consider that their gestures should be understood by a human operator. The reason we asked the participants to begin with *My Gestures* is to see their spontaneous (natural) behavior and avoid biasing their behavior with our three pre-defined gesture sets.

Before completing 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 single mental model gesture sets (*Full Body* and *Puppeteer*). Apart from this short demonstration, the participants received an instruction sheet that showed these gestures (see Figure 3). The participants were encouraged to take as long as they need to study the gestures before proceeding to the task execution.

The verbal description for *Puppeteer* was: *You will control movements of the drone as a puppeteer controls a marionette. This is your neutral position* (Figure 2 was shown). *Imagine that there are two invisible links between your hands and the drone. Movements of the drone depend on your gestures.* The description for *Full Body* was: *The drone will imitate your full body movements. Imagine yourself in place of the drone.* The presentation instruction for *Mixed* was: *In this gesture set to move the drone up and down use the hand movements up and down like the drone is a marionette. To make drone fly right or left extend your arms to the side to imitate an airplane like children usually do it. Tilt your body to the right or left depends on what direction do you want to fly. To fly forward or backward just imagine yourself in place of the drone. Make a step forward to command the drone to fly forward. To rotate the drone, rotate your hand at elbow accordingly.*

As we observed, participants spent no time studying the instruction sheets and started steering the UAV right after the explanation (couple of participants took a few seconds to review gestures from *Mixed* set). During the task, the participant had no access to the lists of gestures.

Study Group	Sequence of study conditions		
	1	2	3
1	*Full Body*	*Puppeteer*	*Mixed*
2	*Puppeteer*	*Mixed*	*Full Body*
3	*Mixed*	*Full Body*	*Puppeteer*
4	*Mixed*	*Puppeteer*	*Full Body*
5	*Full Body*	*Mixed*	*Puppeteer*
6	*Puppeteer*	*Full Body*	*Mixed*

Table 3. The sequence of gesture sets according to balanced Latin square design.

We applied within-subject design to get rid of individual differences. To deal with fatigue or improved performance we counterbalanced the four gesture sets. With 22 participants, four gesture sets, and four routes, participants were assigned to use the gesture sets and routes in different orders. In particular, we used balanced Latin square design to allocate participants to gesture set sequences and to the routes. Latin square designs are often employed in experiments to minimize the number of participants required to detect statistical differences. Generally, potential carryover effects are not balanced out by randomization. Systemic methods are available for equalizing the residual effects [9]. Equal number of times each condition was in different order with other conditions and each condition was followed by different conditions also equal number of times (Table 3).

At first all participants were asked to invent and use their own gesture sets. The other three gesture sets were counter balanced to prevent the problems with the sequence.

In the previous study [17], Peshkova et al. verified that the four routes were equally difficult, but nevertheless we counterbalanced of the four routes as well to make sure that our operator does not know the order of routs. For the first six participants we used Route 1 with *My Gestures*, Route 2 with *Full Body*, Route 3 with *Puppeteer*, and Route 3 with *Mixed*. For the following six participants, shifted the routes by one (Route 4 with *My Gestures*, Route 1 with *Full Body*, Route 2 with *Puppeteer*, and Route 3 with *Mixed*) and etc.

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), time evaluation (how long did it take to finish the route in their opinion) and evaluated statements about just performed gesture set (4-point Likert scale: strongly disagree – disagree – agree – strongly agree). During the entire experiment, the experimenter took notes about think aloud data. After having completed all tasks, the participants judged the four gesture sets regarding their intuitiveness, easiness, and memorability. In the final questionnaire the participants also gave description of gestures and selected their favorite/least favorite gesture set(s). The participants were also asked to explain their choice.

RESULTS

Cognitive Load

We used the following measures to evaluate the level of cognitive load: time perception, dual-task performance, and participants' subjective evaluation [2,3,5,6,8,21,22,24,25]. 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 evaluated the time they spent to complete the task. Block and Gellersen [3] investigated the impact of cognitive load on the perception of time. It has been found that an increase of cognitive load leads to a decrease in time perception [1]. Hart [6] and Zakay and Shub [25] found that participants consistently underestimated time intervals when there was a greater task load: '*increasing task difficulty caused the length of produced intervals to increase*' [2]. The descriptive statistics for the error in time perception for each gesture set is presented in Table 4.

Figure 8 shows the density functions of these time intervals. From these functions we can see that some participants notably overestimated the time spent with *My Gestures*. Thus, we could not conduct one-way repeated measures ANOVA to test the significance of differences (homogeneity assumption is not satisfied) and we had to use non-parametric test. We conducted non-parametric Friedman's test [11] and found that the main effect of gesture set tended to be significant: $\chi^2(3) = 7.25$, $p = 0.06$. Overall, we observed the overestimation of time (Table 4). As we can see, the participants perceived the time spent with *My Gestures* longer than with other gesture sets. The number of participants who underestimated the time was: *My Gestures* – 5, *Full Body* – 7, *Puppeteer* and *Mixed* – 8. That supports (though not significantly) our hypothesis that the cognitive load associated with *intelligent* gesture set (*My Gestures*) was the lowest, with *imitative* (*Full Body*) slightly higher, and with instrumented (*Puppeteer*) – the highest.

Gesture set	Min.	Med.	Mean	Max.	S.D.
My Gestures	-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 4. Descriptive statistics of the error of time perception: Estimated Time – Actual Time.

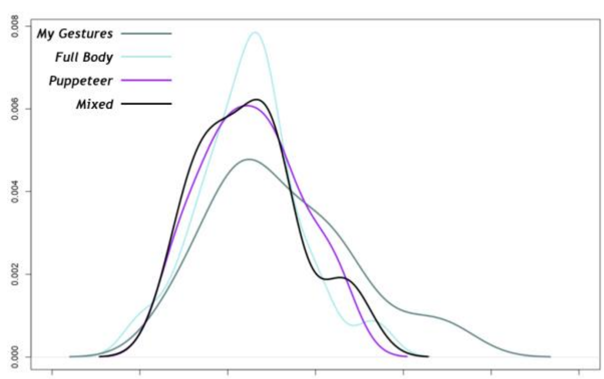

Figure 8. Densities of time deviations.

Khan et al. [8] also investigated the effect of cognitive load on time perception. The authors also reported that higher cognitive load is associated with the higher difference between the actual and estimated time. Based on literature, we computed Directional and Absolute Errors (see Eq. 1 and 2) for a further analysis of participants' time estimation.

$$\text{Directional Error} = \frac{\text{Estimated Time}}{\text{Actual Time}}$$

Equation 1.

$$\text{Absolute Error} = \frac{|\,\text{Estimated Time} - \text{Actual Time}\,|}{\text{Actual Time}}$$

Equation 2.

The descriptive statistics related to these errors are presented in Table 5. For directional errors, the main effect of gesture set, as previously, tended to be significant (Friedman's test: $\chi^2(3)$ = 6.6, p = 0.086). For absolute errors we found no significant difference (Friedman's test: $\chi^2(3)$ = 3.6, p = 0.3). Directional errors are believed to increase with the increase in cognitive load, whereas absolute errors decrease. Though not supported by the results of our statistical tests, *My Gestures* seems to be the easiest to perform.

	Gesture set	Min.	Med.	Mean	Max.	S.D.
Directional Error	My Gestures	0.40	1.39	1.51	3.87	0.77
	Full Body	0.23	1.25	1.21	2.35	0.49
	Puppeteer	0.42	1.21	1.23	2.15	0.50
	Mixed	0.44	1.18	1.19	2.27	0.48
Absolute Error	My Gestures	4.14	54.86	64.32	287.12	65.03
	Full Body	2.33	35.21	41.97	134.99	32.62
	Puppeteer	4.85	38.59	44.05	114.80	32.09
	Mixed	1.61	33.70	39.52	127.17	32.50

Table 5. Descriptive statistics of the directional error.

Dual-Task

We counted how many delays and wrong answers to math questions the participants made while steering the UAV. Figure 9 shows the obtained results. We did not find significant difference between the four gesture sets (Friedman's test: $\chi^2(3)$ = 1.46, p = 0.69).

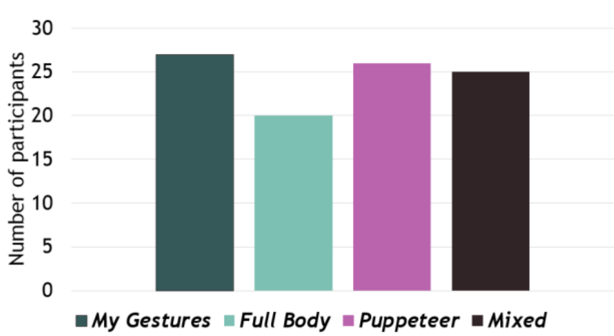

Figure 9. Delays and wrong answers.

Subjective Evaluation

Upon each task completion, our participants evaluated cognitive load experienced with corresponding gesture set using 7-point scale. Figure 10 shows the results of this subjective evaluation. The most frequent evaluation (mode) for *My Gestures* was 3, we assume that it is related to the fact that our participants always started with this 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, that implies that this set was perceived as the most complicated. However, the difference between the four sets was not significant (Friedman's test: $\chi^2(3)$ = 4.38, p = 0.22).

Statements

Upon task completion with each gesture set, the participants responded to five statements (six statements for *My Gestures* only) regarding the used gesture set using a 4-point Likert scale (strongly – disagree – disagree – agree –strongly agree):

S1) I would imagine that most people would learn how to fly the drone with these gestures very quickly.

S2) It was easy to fly the drone and reply to the questions at the same time.

S3) I understand the idea behind the gestures.

S4) The gestures are logically related to each other.

S5) Each gesture individually makes sense for the considered commands.

S6) It was easy for me to come up with gestures.

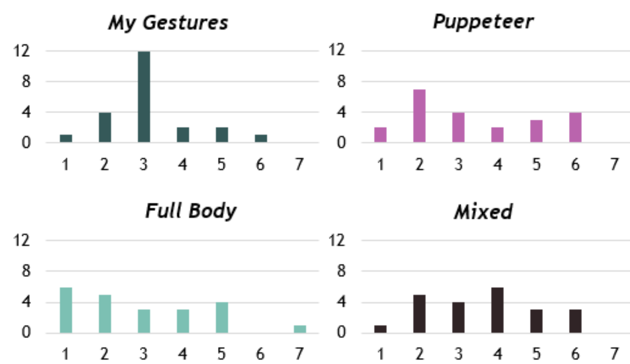

Figure 10. Subjective evaluation of cognitive load: 1 – very low, 7 – very high.

Due to the specificity of gesture-based interaction, we could not identify a standardized questionnaire covering the crucial aspects relevant for our study (intuitiveness, easiness, memorability, and learnability). Thus, because of the absence of established questionnaires for gesture we compiled the statements listed above. We analyzed the differences between the subjective evaluations of the four gesture sets with the Friedman test. For the post-hoc analysis, we applied the Wilcoxon test pairwise with Bonferroni correction.

Figure 11 shows the obtained results with the horizontal axis representing the evaluated statements and the vertical axis indicating the total number of those participants who chose *agree* and *strongly agree* (Note: in the statistical analysis we used all four agreement levels). As all our statements were positive, we can say that the higher the bar the better the evaluation was. Thus, we can conclude that *Mixed* gestures received more negative evaluation in most of the statements compared to *My Gestures*, *Puppeteer* and *Full Body*. S6 concerned only *My Gestures*: 21 participants agreed that it was easy for them to come up with gestures.

Most of the participants agreed that they would imagine that most people would learn how to fly the UAV with *Puppeteer*, *Full Body* and *My Gestures* sets very quickly (S1). In terms of significance, the statistical test revealed that *Full Body* is significantly easier to learn compared to *My Gestures* (S1: $\chi^2(3) = 14.3$, $p = 0.0025$; post-hoc: $p = 0.0075$). Though not visible from Figure 11, most of the participants selected strongly agree with *Full Body* and agree with *My Gestures*. We found no statistical difference between other sets, however, we believe that *Mixed* is actually worse than others: eight participants confused gestures from *Mixed* while steering the UAV; one participant hesitated to show a gesture (*up*) when using *Full Body* and no one confused any gesture from *Puppeteer*.

In all four tasks, the participants found it easy to steer the UAV and reply to math questions at the same time (S2: $\chi^2(3) = 5.75$, $p = 0.12$).

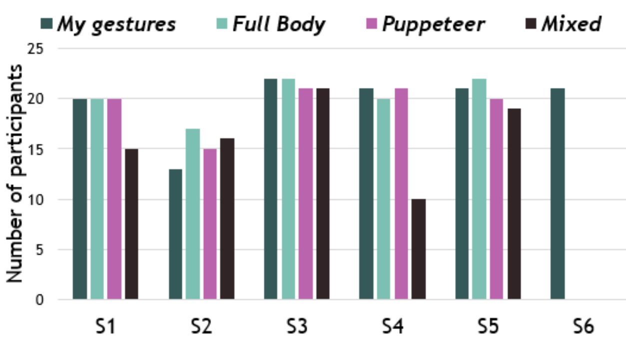

Figure 11. Subjective feedback from questionnaires (after using each gesture set).

Mixed received the worst evaluation in all the statements. The smallest number of the participants understood the idea behind these gestures (S3: $\chi^2(3) = 10.45$, $p = 0.015$) and agreed that the gestures are logically related to each other (S4: $\chi^2(3) = 24.438$, $p < 0.0001$) e.g., 'very confusing'; 'gestures did not seem to be connected to each other'. However, in the first statement the statistical analysis only showed significant difference compared to *Full Body* (S3 [post-hoc]: $p = 0.0067$). Compared to *Mixed*, a significantly greater number of the participants agreed that the gestures from *My Gestures*, *Full Body* and *Puppeteer* were logically related to each other (S4: $\chi^2(3) = 24.438$, $p < 0.0001$; post-hoc: $p = 0.006$, $p = 0.0006$, and $p = 0.0008$, respectively). Also, substantial number of participants agreed, that gestures in *Puppeteer* were more related to each other than in *My Gestures* set (S4: $\chi^2(3) = 24.438$, $p < 0.0001$; post-hoc: $p = 0.0027$). These results confirm that single model gesture sets were indeed perceived by our participants as coherent (logically related to each other).

As expected, we found no significant difference between gesture sets in the fifth statement (S5: $\chi^2(3) = 4.66$, $p = 0.2$). Thus, the participants agreed that all the gestures individually make sense for the considered commands, as opposed to S4 that assessed the overall coherence (internal logical relation between gestures within one set).

Learnability

At the end of the experiment, we asked the participants write down a description of each gesture set for the next participant. Considering the fact that this participant would not see the actual gestures, but control the flight using the written description. We asked participants to either describe each gesture individually or describe the idea behind each set if they think it is sufficient to complete the navigation task (to guess the gestures). The results are presented in Figure 12.

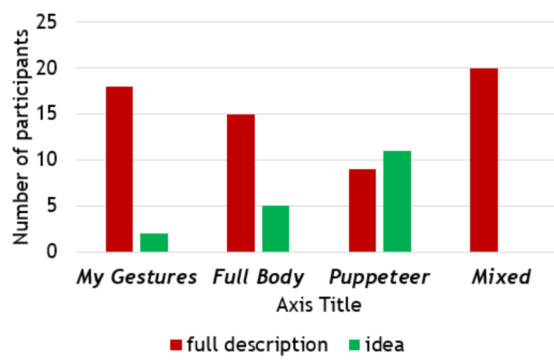

Figure 12. Description of gesture sets.

Five and eleven participants decided that it is enough to give a hint (main idea) to describe *Full Body* (e.g., 'imagine that your body is a drone') and *Puppeteer* set (e.g. 'imagine that you are a puppeteer moving a drone attached to your hands via strings'), respectively. Majority of participants (18 from 22) gave a full description for *My Gestures*. We suppose that they just did not have enough time, or they did not try to recognize the idea behind their own gestures. Though, as we discuss later (see Discussion) there seem to be a common idea behind their behavior (gestures). As expected, all participants gave a full description for *Mixed*.

Priorities

We analyzed the differences between the subjective evaluations of the four gesture sets with the Friedman test. For the post-hoc analysis, we applied the Wilcoxon test pairwise with Bonferroni correction.

In the end of the study, the participants selected their favorite gesture sets and the sets they did not like (multiple selections were possible for both 'favorite' and 'least liked'). Participants also ordered the four gesture sets based on their intuitiveness, easiness, and memorability.

As shown in Figure 13 and Figure 14, *Puppeteer* was favorite of most of the participants while *Mixed* turned out to be the one they did not like. Positive comments regarding *Puppeteer* (e.g., 'easy to remember, easy to show'; 'easy to understand, most intuitive gesture-set'; 'easy to learn and you can concentrate on other tasks') give evidence for their choice. 5 participants mentioned that they got the positive impression from their own gestures (e.g., they liked 'usage of own gestures'; 'coming up with your own gestures, and not having to remember certain once'; 'no need to remember gestures') and *Full Body* (e.g., 'easy to remember'; 'it was like walking – very intuitive'; 'It makes the most sense for me'). 12 participants scored *Mixed Gestures* as leased-liked and augmented 'strange gesture set, not really intuitive'; 'very confusing', 'gestures did not seem to be connected to each other'; 'It wasn't easy to remember because sometimes you had to use your feet and sometimes you had to use your arms'.

Since the participants could select more than one set for both 'favorite' and 'dislike', the sums of the corresponding bars in Figure 13 and Figure 14 are not equal to the total number of the participants – 22.

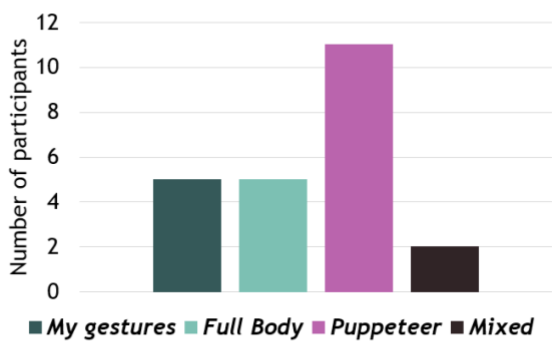

Figure 13. Favorite gesture set.

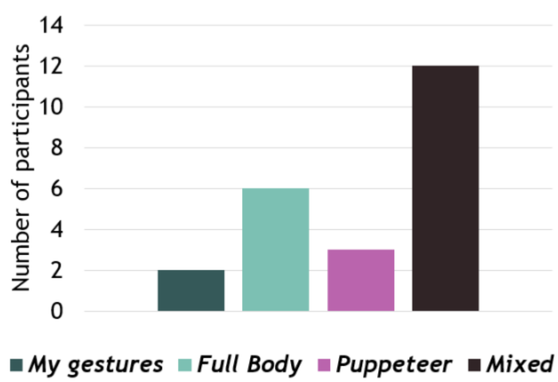

Figure 14. Least-liked gesture set.

We also asked our participants to assign priorities to the four gestures sets in terms of intuitiveness, easiness, and memorability. They could assign the same priority to several sets.

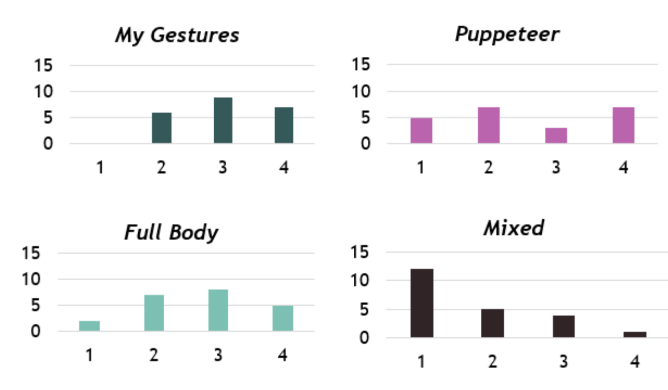

Figure 15. Subjective feedback on intuitiveness: 1 – low intuitiveness, 4 – high intuitiveness.

A significantly greater number of the 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 *My Gestures* and *Mixed* with *Full Body:* $p = 0.0005$, $p = 0.007$, respectively. No significant difference between *Mixed* and *Puppeteer* ($p = 0.058$). The results for intuitiveness are shown in Figure 15.

In terms of easiness, *Mixed* was evaluated significantly more complicated than the other gesture sets: $\chi^2(3) = 17.12$, $p = 0.0007$; post-hoc (*My Gestures*): $p = 0.008$; post-hoc (*Full Body*): $p = 0.0002$; post-hoc (*Puppeteer*): $p = 0.0042$. The results are shown in the Figure 16.

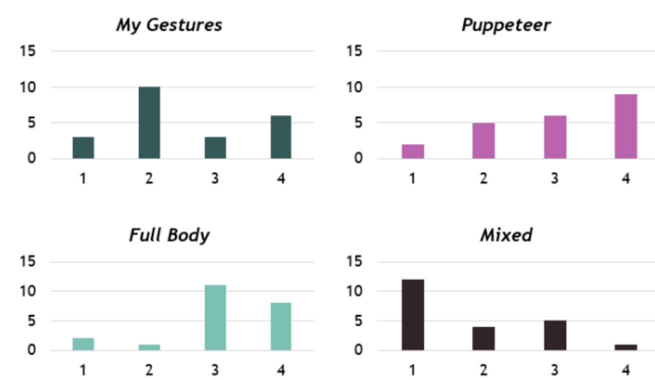

Figure 16. Subjective feedback on easiness: 1 – very easy, 4 – very hard.

In terms of memorability, *Mixed* was evaluated significantly less memorable than the other gesture sets: $\chi^2(3) = 17.08$, p 0.00068; post-hoc (*My Gestures*): $p = 0.02$; post-hoc (*Full Body*): $p = 0.001$; post-hoc (*Puppeteer*): $p = 0.003$. The results are shown in the Figure 17.

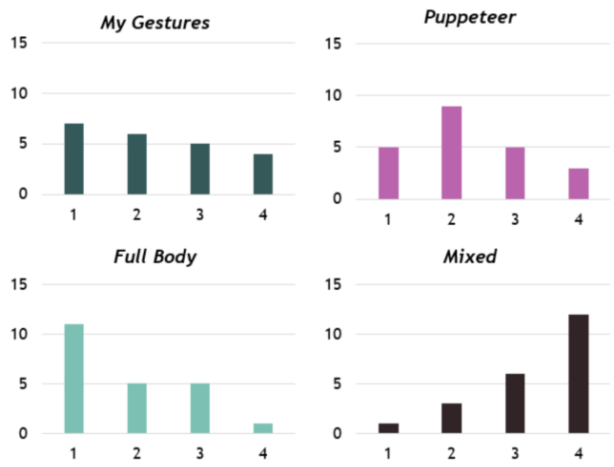

Figure 17. Subjective feedback on memorability: 1 – hard to memorize, 4 – easy to memorize.

Think Aloud Data and Comments

We asked the participants to give an appropriate name for the *Mixed* gesture set. We did not mention that the name of this set is *Mixed*. By letting our participants suggesting their own name for this set, we intended to check if participants would notice and reflect the mixed nature of this set in its name. Figure 18 represents the results as word cloud where words with a bigger size represent names that are more frequent.

Figure 18. Word cloud of names for the *Mixed* gesture set.

Most of participants named *Mixed* set as airplane (7), helicopter (3), and child (2). We suppose that all these names were given according to the most eye-catching ('feature') gesture (airplane: left-right movement and helicopter: rotation) from the set.

Two participants, whose last gesture set was *Mixed*, mentioned that it was 'a combination of everything', 'the most difficult one for me', and 'confusing'.

DISCUSSION

As commonly used in the relevant Human-Computer Interaction works, we selected participants' subjective evaluation and dual-task performance measures to assess cognitive load that different gesture sets impose on users [5,21,22,24,25]. Following the recent research work in Human-Computer Interaction, we also included measures of participants' time perception as an indicator of cognitive load [2,3,6,8].

As a result of our study, we did not find a significant difference 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 *My Gestures* set was associated with the lowest cognitive load among the four sets under study. This finding is interesting considering the fact that all the participants completed their very first navigation task with this gesture set.

Overall, *Mixed* received the worst evaluation compared to the other three sets that supports the previous research [17]. As reported, the difference was not always significant, but it was always inferior to our three single mental model sets. Compared to the other sets *Mixed* is the hardest to learn (see Statements and Learnability subsections): we observed the biggest amount of gesture confusions with this set. Our participants confirmed that this set was not logical and confusing (S4) – it lacks internal coherence. *Mixed* was perceived as the least intuitive, the hardest to use and to memorize. As an outcome, this set was selected by the majority of our participants as the least-liked. Considering that we intentionally selected gestures from different mental models for this gesture set, the obtained result is not surprising, but it does put a stress on the importance of considering adherence to a single mental model when designing gesture-based vocabulary.

Though the obtained results do not support our first hypothesis (cognitive load grows from left to right: *My Gestures* – *Full Body* – *Puppeteer*), we did observe some tendency in favor of this hypothesis. We believe that a potential reason for this negative outcome is poor 'sensitivity' of the employed measures. The difference of cognitive load between the three gesture sets appears to be harder to measure and more precise measures could be considered. Thus, it seems promising to us to further investigate cognitive load associated with different classes of mental models using biofeedback – pupil dilation. We suppose that five statements might not be enough to get the comprehensive data: extended questionnaires could be used in future works. In particular, the first statement could be re-formulated: "I would imagine that a CHILD would learn how to fly the drone with these gestures very quickly". We believe that this statement might help to show the difference between *Full Body* and *Puppeteer*. It would be also interesting to compare a couple of representatives from each class of mental models for a more comprehensive comparison.

In this study, *My Gestures* set represents the *intelligent* class of mental models: our participants were showing gestures that a human operator understands, human-to-human interaction. None of the participants had troubles inventing gestures, no hesitation was observed. We noticed that most of the participants used similar gestures to steer the UAV: they indicated the direction to fly using an index finger (3 participants), one hand (13 participants) or both hands (5 participants) and either rotated their body (7 participants) or rotated their hand at elbow (8 participants) to show the desired rotation direction. From the fact that all our participants immediately come up with similar gestures, we can imply that their behavior was indeed natural.

The participants confirmed that they enjoyed steering the UAV using their own gestures (e.g., 'I did not have to memorize gestures'). Though the participants had a complete freedom to suggest any relevant gestures, we did not find much variety among their behavior. Basically, their gestures could be described via a single sentence: use you hand to indicate the direction to fly or rotate. Thus, the participants tended to follow a single idea and their gestures actually fit into a single mental model that is another interesting finding.

Full Body and *Puppeteer* represent *imitative* and *instrumental* classes of mental models, respectively. As mentioned earlier, we found no significant difference between these gesture sets in terms of cognitive load. Nevertheless, we strongly believe that it makes sense to continue investigating cognitive load associated with these two sets using other more accurate measures that could be biofeedback (e.g., heart-rate variability, brain activity, skin conductance, and eye-tracker) to collect more fine data. In addition, for a more comprehensive investigation, it would be interesting to consider a couple of representative gesture sets from each class of mental models.

Several participants evaluated *Full Body* as physically demanding (e.g., 'too much movement'; 'it might get exhausting'). This finding suggest that *Full Body* is not acceptable for long-term navigation due to its high physical demand, but it might be appropriate for short-term navigation it the field of entertainment [17]. The participants evaluated *Puppeteer* as more appropriate set for a long-term navigation and for people with limited abilities.

Several participants negatively commented individual gestures for the following commands: *rotate left & rotate right* in *Puppeteer* (3 participants found them hard to use) and *Full Body* (1 participant did not like to rotate the whole body as it was not convenient to look at the projection wall); *up & down*: standing on the toes and bending their knees down (due to a physical demand 3 participants did not like it); *left & right* in *Mixed* (6 participants mentioned that they felt silly using this gesture).

CONCLUSION

We conducted a user study, in which we compared four gesture sets in terms of cognitive load as well as intuitiveness, easiness, memorability, and learnability. We investigated whether there is a difference between: 1) the three classes of mental models (*intelligent*, *imitative*, and *instrumented*) and 2) *single mental model* gesture sets and *mixed mental model* gesture sets.

Our findings confirmed our second hypothesis: *mixed mental model* gesture sets are indeed the worst in terms of cognitive load, intuitiveness, easiness, memorability, and learnability.

As for the first hypothesis, our statistical analysis did not confirm it. However, we observed that the representative of the intelligent class of mental models (*My Gestures*) was notably easier in terms of cognitive load, even though our participants always started with this set (no prior experience at all). As for imitative and instrumented classes of mental models, we found no significant difference between *Full Body* and *Puppeteer*. We still believe that there is a difference, but to measure it we need more accurate measures. Thus, we are planning to repeat our study using biofeedback – pupil dilation.

ACKNOWLEDGMENTS

I would like to thank my secondary research supervisor, Ekaterina Peshkova. Without her assistance and dedicated involvement in every step throughout the process, this paper would have never been accomplished. I would like to thank you very much for your support and understanding.

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 *Proceedings 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. 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.
6. 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
7. 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

8. 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.
9. 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.
10. 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
11. 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
12. 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
13. Norman, D. A. (1987). Some observations on mental models. In *Human-Computer Interaction* (241-244). Morgan Kaufmann Publishers Inc. San Francisco, CA, USA.
14. Ekaterina Peshkova. 2017. *Approaching Naturalness and Intuitiveness of Input Vocabularies for UAV Navigation*. Ph.D Dissertation. Alpen-Adria-Universität Klagenfurt, Fakultät für Technische Wissenschaften.
15. 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
16. 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
17. 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
18. 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
19. 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
20. 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
21. 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
22. 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
23. 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 *Proceedings of the 2016 International Conference on Unmanned Aircraft Systems* (ICUAS '16), 1013-1022. https://doi.org/10.1109/icuas.2016.7502665
24. Geert Wernaart. 2013. Cognitive load measurement: Different instruments for different types of load? Master Thesis. Utrecht: Faculty of Social and Behavioral Sciences.
25. 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
26. Questionnaires and other materials used in the conducted user studies: http://www.ias.aau.at/E.Peshkova/
27. [Interactive Systems]. (2018, July 9). User Study. Retrieved from https://www.youtube.com/watch?v=eY8gdGLBYWA&feature=youtu.be