Research Article

Roman Herrmann* and Ludger Schmidt

Design and Evaluation of a Natural User Interface for Piloting an Unmanned Aerial Vehicle

Can gestural, speech interaction and an augmented reality application replace the conventional remote control for an unmanned aerial vehicle?

Abstract: Controlling an unmanned aerial vehicle is challenging and requires an intensive training. One cause is the teleoperation with the conventional input device, the remote control, whose functions are complicate. This paper presents an alternative concept for the teleoperation. Its realization includes a Thalmic Myo gesture control wristlet and a Microsoft HoloLens head-mounted display. These devices are used to implement an augmented reality interface, a tactile feedback and a gesture and speech input. Finally, this implementation has been evaluated with 30 participants and compared with a conventional remote control. The results show that the proposed interface is a good solution but does not reach the performance of the remote control.

Keywords: Natural User Interface, Unmanned Aerial Vehicle, Augmented Reality

1 Introduction

In recent years, unmanned aerial vehicles (UAV) have become increasingly prominent in a wide range of applications. They are used for inspections, delivery or search and rescue operations. Benefits from UAV usage are the possibilities for monitoring activities which are extensive and dangerous for human intervention. At present, the UAV flight control is mainly based on joysticks, buttons and switches. Such devices are not intuitively easy to manipulate, and long training times are needed for precision flying [6, 14]. A conventional remote control (R/C), which is a

*Corresponding author: Roman Herrmann, University of Kassel, Human-Machine Systems Engineering Group, Kassel, Germany, e-mail: R.Herrmann@uni-kassel.de, URL: www.mensch-maschine-systemtechnik.de
Ludger Schmidt, University of Kassel, Human-Machine Systems Engineering Group, Kassel, Germany, e-mail: L.Schmidt@uni-kassel.de, URL: www.mensch-maschine-systemtechnik.de

typical UAV control system, involves the use of several joysticks simultaneously and additional switches for setting the flight parameters and performing other tasks like payload task interactions as well as displays for monitoring the flight and payload parameters (see Fig. 1).

Controlling a UAV is challenging and stressful. The operator has constantly to observe the UAV and its surrounding to avoid collisions. A study has proved that among others the demanding workload is one major issue causing air accidents [36]. To reduce the demanding workload scientific activities focus two aspects: providing assistance for autonomous flight and developing user-friendly interfaces, whose functions are easy to operate. This paper outlines such a user interface.

According to Norman [23] natural user interfaces (NUI) are very popular. Those interfaces are using nonconventional input modalities like gesture, speech or gaze direction. Despite Norman's [23] and Nielsen's [21] warnings about the usability of such interfaces, scientists promise natural and intuitive interaction techniques with „the potential to make human-robot interaction more natural" [26, p. 34]. In addition, the development of low-cost de-

Figure 1: A conventional remote control (R/C) for a professional unmanned aerial vehicle (UAV).

vices such as smart wearables, smart watches or smart glasses, keep pushing researches to envision new interaction possibilities which will replace conventional input devices [15, p. 572]. Following Norman [23] and Nielsen et al. [21] it is important to take account of the application. According to this conclusion, the question arises: Is it possible to replace the conventional input device for UAV with a NUI? To answer this question, we present a NUI composed of an optical see-through head-mounted display (HMD) and a motion control device and compaired this with the conventional input device.

In the following sections, we first present current research results about NUIs (Section 2). Afterwards we discuss current design recommendations (section 3). In section 4, we describe the concept and implementation of a NUI. Then, a laboratory experiment is described (section 5) and the results will be outlined (section 6).

2 Previous Work

The main propose of NUIs is to overcome the constraints of conventional input devices. As Norman [22, p. 210] has noted, the problem with interfaces is, that they are interfaces and that the users must focus attention on them. The underlying cause is the effort to transmit an intention to concrete actions regarding the interface or rather the input device. Conventional input devices, like a R/C, require an enormous effort.

NUI refers to an interface that can easily and efficiently transmit an idea from user's mind into action with little additional effort [16, p. 10]. NUIs are using natural human communications capabilities. Therefore, they should lend themselves to easily learned, intuitive interaction techniques. Among NUIs, gesture-based interfaces have always played a fundamental role in human-machine interaction. Gesture input is a natural communication, because humans are using gestures for communications in their everyday life.

Therefore, researchers have shown increased interest in developing gesture interfaces for controlling UAVs. These developments include definitions of intuitive gestures [26, 2, 15], methods for gesture recognition [7, 18, 32] or the design of feedback [20]. The underlying theory is that a natural and intuitive interface leads to an effective and reliable interface.

These assumptions have been criticized as unproven and too generalized by Nielsen et al. [21] and Norman [23]. First, gestures are not natural per se and second „a gesture interface is not universally the best interface for any application" [21, p. 409]. The first point focuses the lack of

well-known standards for the design of gestural interfaces, which suggests that something like a universal intuitive interaction does not exist. For example, many studies have shown that the cultural background has a vast influence of the intuitive use of gestures [34]. The second point addresses the lack of feedback and accuracy. Also, evaluations of gesture interface for UGVs (unmanned ground vehicles) and UAVs have shown these problems [2, 15]. In previous work we analyzed a gesture set for the teleoperation of a UAV [13]. Based on a laboratory experiment with 14 participants we demonstrated that the gesture-based teleoperation was neither better nor worse than the conventional R/C. But among other things, the participants criticized the missing feedback.

About the second point, the conjunction of a HMD and an augmented reality (AR) interface can compensate the lack of feedback by displaying information in the user's field of view. So far little attention has been paid to the relationship of NUIs, which are using HMDs and gestural input, and the usability related to the teleoperation of UAVs. Hence, Norman's [23] question, whether a NUI is the best interface, can, in case of teleoperation an UAV, not be answered. To close this gap in the current research, a NUI composed of an AR interface and a gesture and speech interface was developed and evaluated.

3 Design Recommendations

Peshkova et al. [26] and Cauchard et al. [2] motivate to use multimodal, intuitive input alphabets like gesture and speech commands. Following Cramar et al. [3], Norman [23] and Raskin [31] it is probable, that a universal intuitive gesture or speech command alphabet for the teleoperation of UAVs does not exist. However, gesture or speech alphabets, which are corresponding to mental models of target users, are felt intuitive [26, p. 40].

In addition, there are other requirements focusing ergonomic [21, 19], technical [35, 33] or usability (like feedback or execution times) aspects. For example, Nielsen et al. [21, p. 414] disclose six principles in ergonomics, including the avoidance of non-ergonomic positions. According to McMillian [19] gestures should be concise and quick to minimize fatigue. To ensure learning and performance advantages gestures should be designed in such a way, that natural co-ordination of the body can be employed to co-ordinate multiple degrees of freedom to the external device [19].

Certainly, these diverse requirements are not necessarily compatible with each other. For example, the require-

ment on the user not to wear markers, gloves or not to fix the background conflicts with the requirement of accuracy (detection, tracking and recognition) [35]. In fact, balancing these conflicting factors according to ensure a successful teleoperation of a UAV may be the key to a usable NUI [37, p. 162].

There can be no doubt that feedback is one important aspect to ensure a usable interface [5, 23, 21]. To avoid limitations related to a constant line-of-sight between operator and UAV, visual information should be displayed in the field of view but not overlap the UAV. One possibility to fulfil these requirements is the usage of a HMD with an AR interface. Concerning displaying information (absolute or relative position, colors) ISO 9241-13 suggests, that status information (like battery status or signal strength) should be displayed in a consistent way, always placed at the same position. Furthermore, information should be displayed noticeable for users, but users should not be disturbed in any case and be distracted as less as possible.

As Livingston [17, p. 8] has noted, there are only a few studies on the human factors associated with AR systems. Often studies focused on the use of HMD in aerospace applications. In a research report from the NASA researches identified "attention" or "cognitive capture" as one major issue causing air accidents [28]. This term describes the issue that it is difficult for pilots to switch between information displayed in a HMD and information in the "outside world". Furthermore, efficient text visualization in AR displays is critical because it is sensitive to display technology, ambient illumination and various backgrounds [4].

Based on previous work [4, 8] we experimented the effects on the readability during a UAV-teleoperation of three different text styles (plain text and bill boarded text) [12]. To evaluate the different text styles 30 participants were asked to pilot a UAV through a course, while they wore a HMD. At randomized intervals, textual information were displayed in the HMD which the participants had to react. Our findings can be summarized as follows: Textual infor-

mation should use font colors black or red on a white billboard color.

However, it cannot be assumed that the user is always focusing fade-in textual information in the HMD. It is more likely to be assumed that the focus is on the UAV and its surroundings. To achieve an effective and efficient feedback ISO 9241-13, Norman [23], Zhai et al. [37] propose the usage of multimodal feedback.

4 Concept and Implementation

NUIs and AR visualization in HMDs date back over a long time [24], but so far implementations have been bulky and not very user-friendly. The technological development for HMDs and wearables has already achieved considerable progress, so it is possible to design and implement a NUI in compliance with the requirements formulated in section 3. Regarding design recommendations, we designed and implemented a NUI.

4.1 Gesture-Based Interface

A key criterion for the design of the gesture-based interface was a high usability. To realize this criterion we prioritized intuitiveness, short execution times and ergonomics aspects. In previous work we defined a gesture set for the teleoperation of a UAV [13]. This gesture set is based on the mental model, that the forearm imitates the attitude of the UAV. In detail the elbow joint (pronation, supination and flexion) specifies the flight attitude (see Fig. 2). A tentative conclusion at this point would be that the user probably possess or develops a similar model. In addition, this gesture set allows short execution times.

This gesture set should offer learning and performance advantages, because the natural co-ordination of the forearm can be employed to co-ordinate the attitude of the UAV. In line with Nielsen's [21] ergonomic require-
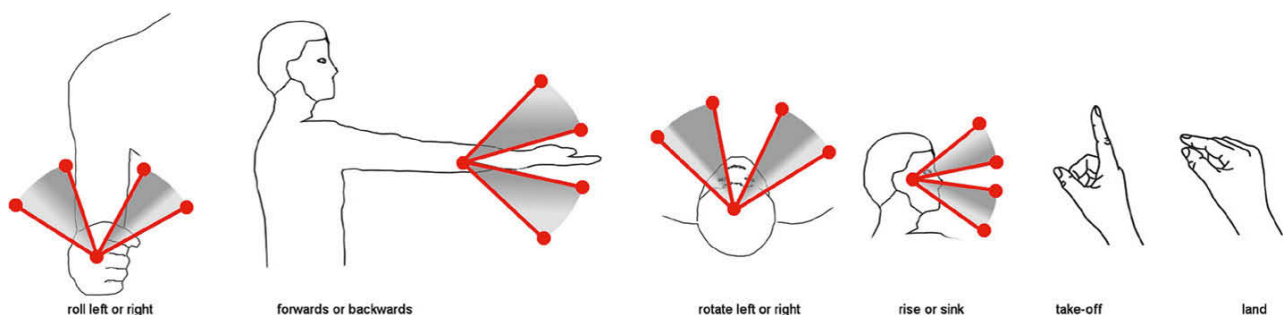

Figure 2: Set of gesture commands. Red lines represent thresholds.

ments we defined relaxed neutral positions and avoid non-ergonomic positions. Therefore, we used the previous work of Bleyer et al. [1]. As a result, the control concept is based on threshold values (represented by red lines in Fig. 2). For example, to command a UAV to fulfil a left movement, a threshold must be surpassed. This threshold represents a certain pronation angle and is based on an ergonomically favorable position. Depending on how much the threshold has been surpassed the UAV moves faster.

This interaction concept can be interpreted as a direct mapping of the operator's movements to the UAV motion. The difficulty is to design gestures for controlling "up" and "down" as well as "rotate left" and "rotate right" which fit in this mental model and are ergonomic. Research by several scholars [25, 6, 27] suggests abduction or adduction of the arms. In order to realize short execution times, we reject this approach. Instead, we decided to use human head motions to control these functions as well as mentioned by Higuchi and Rekimoto [14] and Hegenberg et al. [9]. The resulting human gesture commands are illustrated in Fig. 2.

4.2 Speech-Based Interface

Natural interaction includes gestures as well as speech. Therefore a NUI should not be limited to gestures. In our previous work we observed that users often use gestures in combination with the appropriate speech command, e. g. while preforming a "roll left" gesture and saying "left". It might therefore be advantageous to implement a speech-based interface in order to achieve a high usability. In recent studies, Peshkova et al. [26] and Cauchard et al. [2] have suggested several speech commands, which we used (see Tab. 1). The speech commands are to be interpreted as follows: execute the command for two seconds if no other speech command is given. While a speech command is executed no gesture command is accepted.

4.3 Feedback

As pointed out in section 3 we designed three different areas for the visual feedback. The central area of the field of view is kept free to allow a direct line of sight to the UAV without occlusions. Below this area we designed two areas, one to display alerts, like low battery, and another one to display a constant visualization of the UAV attitude. The visualization of status information, like battery or received signal strength, is displayed above the central area. Warning messages and the constant information about UAV attitude are visualized in font colors red or black on a white

Table 1: Set of speech commands.

UAV activity	Speech command
take off	start, starte (German), starten (German), hebe ab (German)
land	land, lande (German), landen (German)
position hold	stop, halte (German), schwebe (German), Stopp (German), Halt (German)
roll left	left, links (German), fliege links (German)
roll right	right, rechts (German), fliege rechts (German)
forwards	for, vor (German), vorwärts (German), fliege vorwärts (German)
backwards	back, zurück (German), fliege zurück (German)
up	up, steige (German), steigen (German), hoch (German)
down	down, sinke (German), sinken (German), runter (German)
rotate left	rotate left, rotiere links (German), giere links (German)
rotate right	rotate right, rotiere rechts (German), giere rechts (German)

billboard color, whereas status information are displayed in font color white without billboard (see Fig. 3). Further we implemented audible feedback as soon as the UAV is changing its attitude. For this purpose, we used the google voice response to create short verbal prompts. In addition, we have opted for tactile feedback, provided when the thresholds (red lines in Fig. 2) are exceeded.

4.4 Hardware Components

For the recognition of the forearm pose we used the Myo wristlet by Thalmic Labs. It is a low-cost 8-channel, consumer-grade surface electromyography (sEMG) device. The Myo wristlet is non-invasive, as it enables users to simply put it on without any preparation. This ensures that the requirement "come as you are" has almost been fulfilled [35, p. 63]. In addition, the wristlet includes a 9-axis inertial measurement unit (IMU). The IMU can be used to determine the pitch, roll and yaw of the user's forearm. Moreover, it provides tactile feedback by vibration.

Several companies are researching and developing hardware in form of glasses to make AR-applications possible. One resulting device is the Microsoft HoloLens. It consists of depth camera, speakers, holographic lenses and an IMU, composed of accelerometer, gyroscope and magnetometer.

We used an unmodified multi-rotor Parrot Bebop UAV. The device weighs approximately 400 g with a maximum flight time of 11 min and can be controlled via a mobile device, or via an imitated R/C named Skycontroller.
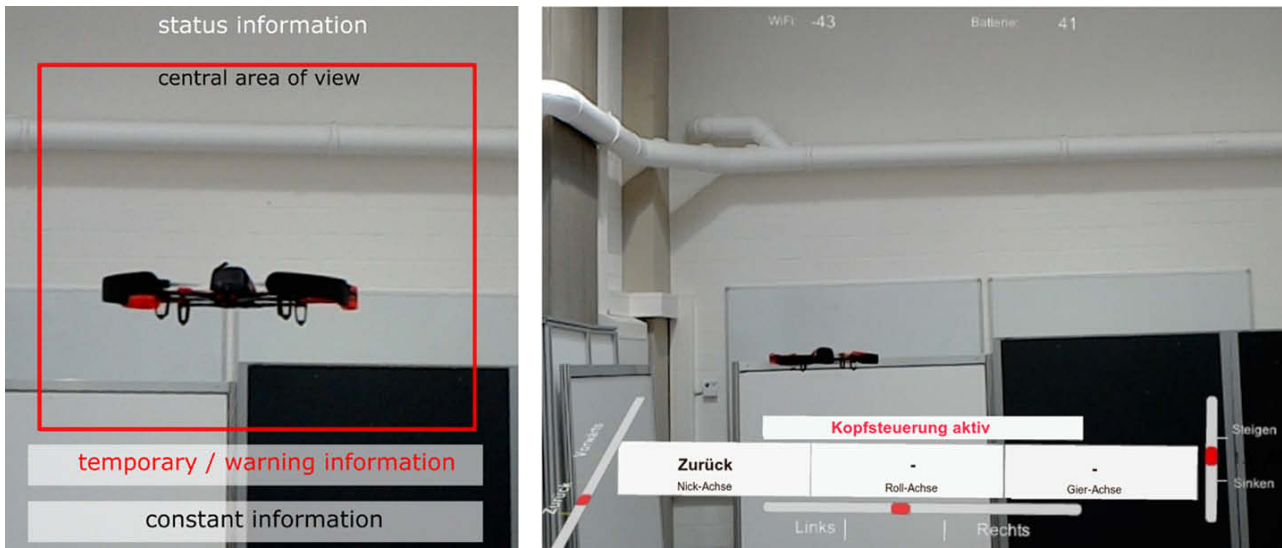
Figure 3: Concept of the AR interface (left) and its realization (right).

4.5 Architecture of the System

We used Robot Operation System (ROS), a hybrid P2P distributed architecture, to pass information [30] and assigned the *ROS Master* role to a laptop (see Fig. 4). The UAV and the Myo are communicating with the *ROS Master* through a wireless network. We implemented three nodes, which are preforming computation. The first node is used to get the current pose of the Myo and to instruct it to vibrate (*Myo Node in Fig. 4*). The second node is used to con-

trol the UAV (*UAV Node in Fig. 4*). The third node is collecting and processing information from the *Myo* and *UAV Node* (*Control Node in Fig. 4*) and passing the information to the main control, which is not part of the ROS architecture as well as the Microsoft HoloLens. Instead, we implemented the main control (*Main Control in Fig. 4*), which is used to control the whole process, as a stand-alone program and a mixed reality application for the HoloLens, which communicates with the other elements of the architecture via User Datagram Protocol (UDP). *ROS Mas-*
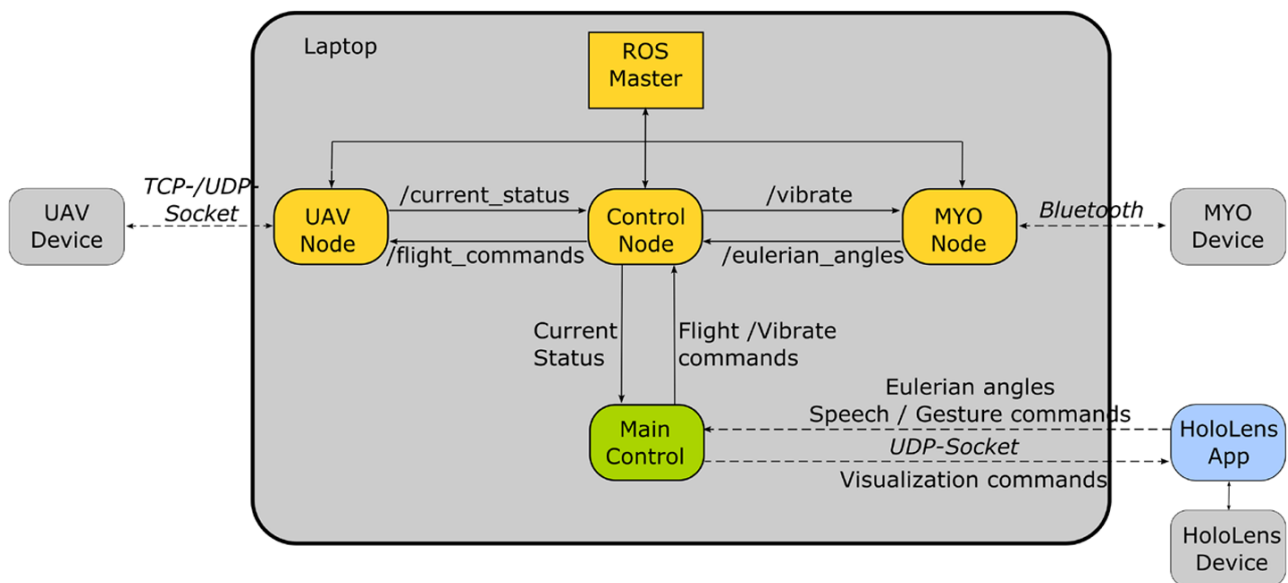

Figure 4: Architecture of the system: Elements of the ROS infrastructure (yellow), main control (green), mixed reality application for the HoloLens (blue), hardware components (grey). Dotted arrows are representative of transmission protocol.

ter, UAV Node, Control Node, Myo Node and *Main Control* are running on the same laptop. The *Main Control* contains the control algorithms to determinate the state variables, which indicates the desired status of the flight attitude, the visualizations in the HoloLens and the vibrate behavior of the Myo. According to the state variables the *Main Control* is passing appropriate orders to the *Control Node* and the mixed reality application for the HoloLens (*HoloLens App in Fig. 4*). The HoloLens application was developed with Unity Personal and C# in combination with the MixedRealityToolkit-Unity (for speech recognition) and runs on the HoloLens, whereas the *UAV Node*, *Control Node* and *Main Control* were developed in C / C++ and the *Myo Node* in Python. For the development of the *UAV Node* we used the Software Development Kit (SDK) provided by Parrot and for the *Myo Node* the Myo SDK.

5 Method

In order to evaluate the usability of the NUI, a laboratory experiment with 30 participants was conducted. Within this experiment the Parrot Bebop UAV had to be piloted with two modalities of control (Skycontroller and NUI). To avoid learning and "carryover" effects we decided to split the participants into two groups. 15 participants (6 female) with a mean age of 30 ± 3.25 years were using the NUI. 15 participants (7 female) with a mean age of 31 ± 4.73 years were using the Skycontroller. Within the NUI-group 8 participants had used a gesture control before (in the context of games) and within the Skycontroller-group 4 participants had used an R/C to teleoperate a UAV.

In the beginning of the experiment, participants were informed about procedure of the experiment and filled out a demographic questionnaire. To evaluate the interfaces the participants were asked to pilot the Parrot Bebop UAV through four different courses (see Fig. 5 and Fig. 6).


Figure 6: Participant equipped with HoloLens and Myo piloting the UAV through course 3.

Within these courses different flight maneuvers had to be executed. The aims were to take off, bypass obstacles and to land within a mark (landing area). Course 1 required to bypass an obstacle by rising. Course 2 required to fulfil left and right roll maneuvers. Course 3 and 4 included a right or rather left rotation.

Course 1 and course 2 had to be completed three times, whereas course 3 and 4 two times. The courses were considered as accomplished successfully, if the investigator did not have to intervene to avoid UAV collisions, the UAV did not crash and the UAV was landed within the marks.

As dependent variables, effectivity was measured by successful executions, efficiency measured by the required time for passing (from the take off to landed UAV), and suitability for learning determined by the questionnaire ISONORM 9241/10 [29]. The latter questionnaire operationalizes the ergonomic principle "suitability for learning". For this principle, five bi-polar statements are to be rated on a 7-point scale ranging from "- - -" (1) to "+++" (7). Following this metric, we created a further questionnaire, which focuses the satisfaction related to the modality of control. The questionnaire required to rate five bi-polar statements with the same rating scale as the ISONORM 9241/10. After completing all four courses a structured interview with questions about the interface was conducted. In the end, the participants had the opportunity to men-
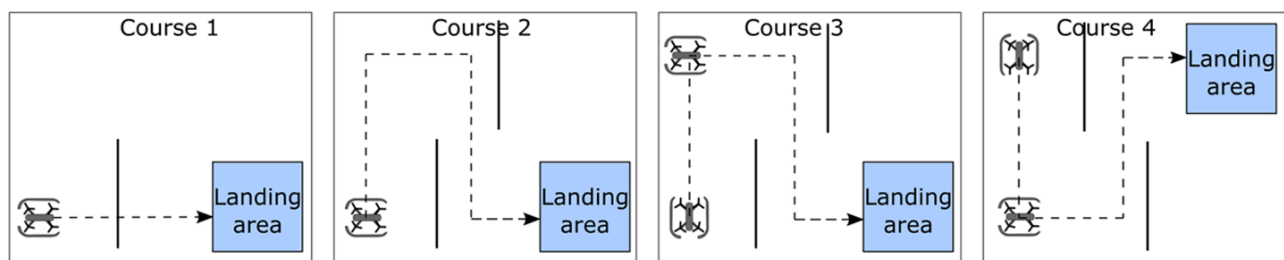
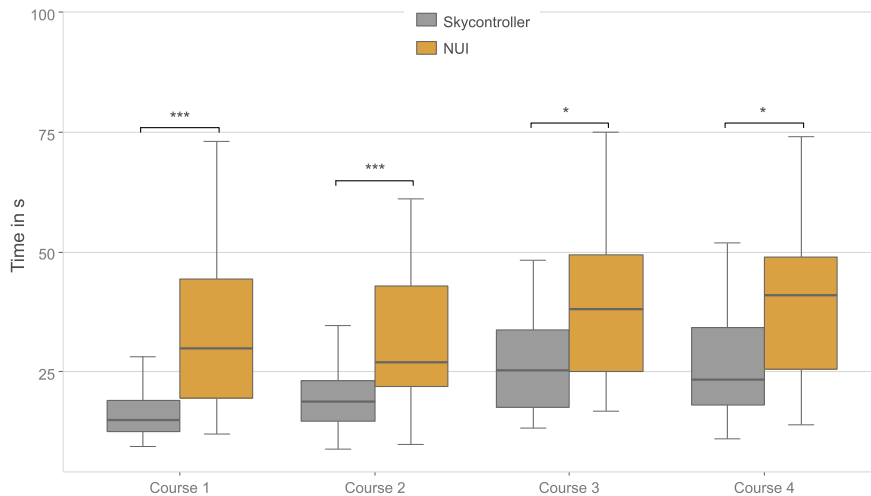**Figure 5:** Visualization of the four courses.

Figure 7: Time to complete the courses.

tion non-addressed aspects and general comments and suggestions.

6 Results

6.1 Effectivity

140 of 150 flights were successfully completed using the NUI, whereas 132 of 150 flights were successfully executed using the Skycontroller. A chi-square test of independence was performed to examine the relation between type of control (NUI or Skycontroller) and successful completed tasks. The relation between these variables was not significant, $\chi^2(1) = 2.52$, $p = .11$. Tab. 2 shows the results in detail.

Regarding the NUI, 2671 gesture and 206 speech commands were used (see Tab. 3). Most speech commands

Table 2: Effectivity comparing both modalities of control (Skycontroller vs. NUI).

Successful completions with	Course 1	Course 2	Course 3	Course 4
Skycontroller	41 of 45	40 of 45	25 of 30	26 of 30
NUI	43 of 45	43 of 45	27 of 30	27 of 30

Table 3: Used speech and gesture commands (NUI, only successful completed tasks regarded).

Command	N	Course 1	Course 2	Course 3	Course 4
Gesture	2671	695	643	655	678
Speech	206	33	59	56	58

were used for "take off", "land", "rotate left" and "rotate right".

6.2 Efficiency

The time taken to complete the courses was significantly shorter with the Skycontroller (see Fig. 7 and Tab. 4). Extremes values were excluded from further analysis.

Table 4: Efficiency comparing both modalities of control (Skycontroller vs. NUI).

Course	Modality of control	N	M [s]	SD [s]	Test
Course 1	Skycontroller	41	16.34	4.90	$t(50.55) = -6.71^a$
	NUI	43	33.41	15.89	$p < .001$, $r = .69$
Course 2	Skycontroller	40	19.92	7.50	$t(63.47) = -5.26^a$
	NUI	43	33.27	14.69	$p < .001$; $r = .55$
Course 3	Skycontroller	26	29.49	14.76	$t(51) = -2.25^a$
	NUI	27	40.81	21.21	$p = .029$; $r = .30$;
Course 4	Skycontroller	25	27.49	11.78	$t(50) = -2.91^a$
	NUI	27	39.25	19.26	$p = .011$; $r = .38$

[a] Normal distribution, homogeneity of variance, independent t-test.

We observed two significant learning effects. On average, participants required more time of completing the first-time course 2 and 4, using the NUI, than for last-time completion (see Tab. 5). Besides that, no significant results regarding learning effects were found (see Tab. 6).

6.3 Questionnaires and Interviews

A Mann-Whitney test indicated that the suitability for learning was not greater for the Skycontroller ($M = 4.9$)

Table 5: Efficiency comparing time taken to complete in different executions (NUI).

Course	Execution	N	M [s]	SD [s]	Test
Course 1	First	13	39.78	14.89	$U = 25.5, z = -1.70$[b]
	Last	13	30.21	17.12	$p = .09$
Course 2	First	13	36.42	14.05	$t(12) = 2.92$[a]
	Last	13	25.21	8.69	$p = .013$
Course 3	First	11	46.00	26.18	$t(11) = 2.07$[a]
	Last	11	35.16	15.95	$p = .062$
Course 4	First	10	43.26	22.40	$t(11) = 3.77$[a]
	Last	10	34.27	16.35	$p = .003$

[a] Normal distribution, homogeneity of variance, dependent t-test.
[b] One or both distributions not normal, Mann-Whitney test.

Table 6: Efficiency comparing time taken to complete in different executions (Skycontroller).

Course	Execution	N	M [s]	SD [s]	Test
Course 1	First	14	18.19	6.21	$t(12) = 1.25$[a]
	Last	14	15.49	4.05	$p = .23$
Course 2	First	13	21.22	7.70	$t(12) = 0.19$[a]
	Last	13	20.54	9.62	$p = .85$
Course 3	First	12	30.52	16.45	$U = 33, z = .00$[b]
	Last	12	29.69	16.03	$p = 1.00$
Course 4	First	12	24.51	11.86	$t(9) = -.84$[a]
	Last	12	29.08	12.62	$p = .42$

[a] Normal distribution, homogeneity of variance, dependent t-test.
[b] One or both distributions not normal, Mann-Whitney.

than for the NUI ($M = 5.2$), $U = 88.50$, $p = .325$. However, a Mann-Whitney test indicated that the satisfaction was greater for the NUI ($M = 5.4$) than for the Skycontroller ($M = 4.5$), $U = 14$, $p < .001$.

The participants emphasized that they had fun using the NUI and that it requires only a few minutes to learn the NUI. All participants, who used the Skycontroller, stated that the Skycontroller allows a precise control. 90 % of the participants said, that the NUI was intuitive, whereas only 27 % stated, that the Skycontroller was intuitive.

Some of the NUI participants complained about the head motions to control the rotation and the altitude. Often using head motions resulted in a lost line of sight to the UAV. In addition, they mentioned, that they had problems to coordinate head and hand movements. One third (30 %) of the participants stated that they did not pay attention to the textual information inside the HMD. Instead they paid attention to the vibrations and the verbal prompts.

The NUI participants were asked to suggest gestures for controlling "up" and "down", "rotate left" and "rotate right", which are intuitive for them. The suggested gestures were very different, including different movements of arms, head or body.

7 Discussion

In our study, which we partly presented on a conference [10], a NUI for the teleoperation of a UAV has been developed and evaluated. This NUI has been compared to a conventional input device. Regarding the effectivity it is not possible to prefer one modality of control. Our findings support the view that a conventional input device is more efficient than the NUI. This result must be interpreted with caution. Several participants noted that there were too many possibilities (speech and gestures commands) to pilot the UAV. Thus, participants considered at length whether which possibility to choose. Therefore these results must be treated as tentative as more research is conducted to identify the relation between multimodal, intuitive input alphabets and efficiency. Partially we could observe significant learning effects with the NUI. This should be considered in further experiments.

The satisfaction was greater for the NUI than for the Skycontroller and participants confirmed that the NUI is intuitive. Our findings strongly support the fact that an intuitive interface is not the most efficient interface. The answer to the question whether a NUI can replace the conventional remote control depends on the relevance of efficiency. Future research should concentrate on improving efficiency.

This overall impression could also be confirmed at a demonstration at a conference where layperson people were able to pilot a UAV (see Fig. 8) [11].

Figure 8: Demonstration of the NUI at a conference. Laypersons were able to pilot the UAV.

References

[1] Bleyer, T., Hold, U., Rademacher, U., & Windel, A. (2009). Belastungen des Hand-Arm-Systems als Grundlage einer ergonomischen Produktbewertung – Fallbeispiel Schaufeln. 1. Auflage. Dortmund: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin.

[2] Cauchard, J. R., Zhai, K. Y., Spadafora, M., & Landay, J. A. (2016). Emotion Encoding in Human-Drone Interaction. In *11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 263–270). Piscataway, NJ, USA: IEEE Press.

[3] Cramar, L., Hegenberg, J., & Schmidt, L. (2012). Ansatz zur experimentellen Ermittlung von Gesten zur Steuerung eines mobilen Roboters. In VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik, *Useware 2012: Mensch-Maschine-Interaktion (Kaiserslautern 2012)* (Vols. VDI-Berichte 2179, pp. 173–183). Düsseldorf: VDI-Verlag.

[4] Debernardis, S., Fiorentino, M., Gattullo, M., Monno, G., & Uva, A. E. (2014). Text readability in head-worn displays: color and style optimization in video versus optical see-through devices. *IEEE transactions on visualization and computer graphics, 20*(1), pp. 125–139.

[5] DIN EN ISO 9241-13. (2011). Ergonomie der Mensch-System-Interaktion – Teil 13: Benutzerführung.

[6] Dong, M., Cao, L., Zhang, D.-M., & Guo, R. (2016). UAV flight controlling based on Kinect for Windows v2. In *International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)* (pp. 735–739). IEEE.

[7] Elmezain, M., Al-Hamadi, A., Appenrodt, J., & Michaelis, B. (2008). A Hidden Markov Model-based continuous gesture recognition system for hand motion trajectory. In *19th International Conference on Pattern Recognition* (pp. 1–4). Piscataway, NJ: IEEE.

[8] Gabbard, J. L., Swan, J. E., & Hix, D. (2006). The Effects of Text Drawing Styles, Background Textures, and Natural Lighting on Text Legibility in Outdoor Augmented Reality. *Presence: Teleoperators and virtual environments, 15*(1), pp. 16–32.

[9] Hegenberg, J., Cramar, L., & Schmidt, L. (2012). Task- and User-Centered Design of a Human-Robot System for Gas Leak Detection: From Requirements Analysis to Prototypical Realization. In I. Petrovic, & P. Korondi, *10th International IFAC Symposium on Robot Control (Dubrovnik 2012)* (pp. 793–798). Dubrovnik: IFAC.

[10] Herrmann, R., & Schmidt, L. (2017). Gestaltung und Evaluation einer natürlichen Flugrobotersteuerung. In M. Burghardt, R. Wimmer, C. Wolff, & C. Womser-Hacker, *Mensch und Computer 2017 – Tagungsband (Regensburg 2017)* (pp. 147–158). Bonn: Gesellschaft für Informatik e. V.

[11] Herrmann, R., & Schmidt, L. (2017). Natürliche Benutzungsschnittstelle zur Steuerung eines Flugroboters. In M. Burghardt, R. Wimmer, C. Wolff, & C. Womser-Hacker, *Mensch und Computer 2017 – Workshopband (Regensburg 2017)* (pp. 637–640). Bonn: Gesellschaft für Informatik e. V.

[12] Herrmann, R., Hegenberg, J., & Schmidt, L. (2016). Evaluation des Leitstands eines Boden-Luft-Servicerobotiksystems für eine Produktionsumgebung. In VDI Wissensforum GmbH, *Useware 2016* (pp. 187–200). Düsseldorf: VDI Verlag GmbH.

[13] Herrmann, R., Hegenberg, J., Ziegner, D., & Schmidt, L. (2016). Empirische Evaluation von Steuerungsarten für Flugroboter. In Gesellschaft für Arbeitswissenschaft e. V., *Arbeit in komplexen Systemen – Digital, vernetzt, human?! 62. Kongress der Gesellschaft für Arbeitswissenschaft (Aachen 2016)* (pp. 1–6 (A.4.9)). Dortmund: GfA-Press.

[14] Higuchi, K., & Rekimoto, J. (2013). Flying head: a head motion synchronization mechanism for unmanned aerial vehicle control. In W. E. Mackay, *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (pp. 2029–2038). New York, NY: ACM.

[15] Jones, G., Berthouze, N., Bielski, R., & Julier, S. (2010). Towards a situated, multimodal interface for multiple UAV control. In *2010 IEEE International Conference on Robotics and Automation* (pp. 1739–1744). Piscataway, NJ: IEEE.

[16] Lee, J. C. (2010). In search of a natural gesture. *XRDS: Crossroads, The ACM Magazine for Students, 16*(4), pp. 9–13.

[17] Livingston, M. A. (2013). Issues in Human Factors Evaluations of Augmented Reality Systems. In W. Huang, L. Alem, & M. A. Livingston, *Human factors in augmented reality environments* (pp. 3–9). New York: Springer.

[18] Mäntylä, V.-M. (2001). Discrete hidden Markov models with application to isolated user-dependent hand gesture recognition. pp. 2–104.

[19] McMillan, G. R. (1998). The technology and applications of gesture-based control. In T. R. Anderson, G. McMillian, J. Borah, & G. M. Rood, *Alternative Control Technologies, Human Factors Issues* (pp. 1–11). Canada Communication Group.

[20] Monajjemi, M., Mohaimenianpour, S., & Vaughan, R. (2016). UAV, come to me: End-to-end, multi-scale situated HRI with an uninstrumented human and a distant UAV. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 4410–4417). IEEE.

[21] Nielsen, M., Störring, M., Moeslund, T., & Granum, E. (2004). A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for HCI. In A. Camurri, & G. Volpe, *Gesture-Based Communication in Human-Computer Interaction* (Vol. 2915, pp. 409–420). Springer Berlin Heidelberg.

[22] Norman, D. A. (1990). Why interfaces don't work. In B. Laurel, & S. J. Mountford, *The Art of human-computer interface design* (pp. 209–219). Addison-Wesley.

[23] Norman, D. A. (2010). Natural User Interfaces Are Not Natural. *interactions, 17*(3), pp. 6–10.

[24] Oehme, O., Wiedenmaier, S., Schmidt, L., & Luczak, H. (2001). Empirical Studies on an Augmented Reality User Interface for a Head Based Virtual Retinal Display. In M. J. Smith, & G. Salvendy, *Systems, Social and Internationalization Design Aspects of Human-Computer Interaction: Proceedings of the HCI International 2001* (pp. 1026–1030). Mahwah: Erlbaum.

[25] Peshkova, E., Hitz, M., & Ahlström, D. (2016). Exploring User-Defined Gestures and Voice Commands to Control an Unmanned Aerial Vehicle. In R. Poppe, J.-J. Meyer, R. Veltkamp, & M. Dastani, *Intelligent Technologies for Interactive Entertainment* (pp. 47–62). Cham: Springer International Publishing; Imprint: Springer.

[26] Peshkova, E., Hitz, M., & Kaufmann, B. (2017). Natural Interaction Techniques for an Unmanned Aerial Vehicle System. *IEEE Pervasive Computing, 16*(1), pp. 34–42.

[27] Pfeil, K., Koh, S. L., & LaViola, J. (2013). Exploring 3d gesture metaphors for interaction with unmanned aerial vehicles. In J. Kim, J. Nichols, & P. Szekely, *Proceedings of the 2013 international conference on Intelligent user interfaces* (pp. 257–266). New York, NY: ACM.

[28] Prinzel, L. J., & Risser, M. (2004). *Head-Up Displays and Attention Capture*. Springfield: National Technical Information Service.

[29] Prümper, J. (1997). Der Benutzungsfragebogen ISONORM 9241/10: Ergebnisse zur Reliabilität und Validität. In R. Liskowsky, B. M. Velichkovsky, & W. Wünschmann, *Software-Ergonomie '97: Usability Engineering: Integration von Mensch-Computer-Interaktion und Software-Entwicklung* (pp. 254–262). Stuttgart: B. G. Teubner.

[30] Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., Ng, A. Y. (2009). ROS: an open-source Robot Operating System.

[31] Raskin, J. (1997). Intuitive equals Familiar. *Communications of the ACM, 37*(9), pp. 17–18.

[32] Schlenzig, J., Hunter, E., & Jain, R. (1994). Recursive identification of gesture inputs using hidden markov models. In *Proceedings of the Second IEEE Workshop on Applications of Computer Vision* (pp. 187–194). IEEE.

[33] Schmidt, L., Herrmann, R., Hegenberg, J., & Cramar, L. (2014). Evaluation einer 3-D-Gestensteuerung für einen mobilen Serviceroboter. *Zeitschrift für Arbeitswissenschaft, 68*(3), pp. 129–134.

[34] Urakami, J. (2014). Cross-cultural comparison of hand gestures of Japanese and Germans for tabletop systems. *Computers in Human Behavior, 40*(0), pp. 180–189.

[35] Wachs, J. P., Kölsch, M., Stern, H., & Edan, Y. (2011). Vision-based hand-gesture applications. *Communications of the ACM, 54*(2), pp. 60–71.

[36] Williams, K. W. (2004, Dec). A Summary of Unmannded Aircraft Accidents/Incident Data: Human Factors Implications. (F. A. Dept. Transportation, Ed.) Final Report DOT/FAA/AM-04/24.

[37] Zhai, S., Kristensson, P. O., Appert, C., Anderson, T. H., & Cao, X. (2012). Foundational issues in touch-surface stroke gesture design—an integrative review. *Foundations and Trends® in Human Computer Interaction, 5*(2), pp. 97–205.

Bionotes

Roman Herrmann
University of Kassel, Human-Machine Systems Engineering Group, Kassel, Germany
R.Herrmann@uni-kassel.de

Roman Herrmann is a PhD candidate in the Human-Machine Systems Engineering Group in the Department of Mechanical Engineering at the University of Kassel. His research interests include human-robot interaction and user-oriented interface design. He received a MSc in computer science at the University of Paderborn.

Ludger Schmidt
University of Kassel, Human-Machine Systems Engineering Group, Kassel, Germany
L.Schmidt@uni-kassel.de

Univ.-Prof. Dr.-Ing. Ludger Schmidt has studied electrical engineering at the RWTH Aachen University in Germany. There he also worked as a research assistant, research team leader, and chief engineer at the Institute of Industrial Engineering and Ergonomics. Afterwards he was the head of the department "Ergonomics and Human-Machine Systems" at today's Fraunhofer Institute for Communication, Information Processing and Ergonomics in Wachtberg near Bonn. In 2008, he became Professor of Human-Machine Systems Engineering in the Department of Mechanical Engineering at the University of Kassel. He is director of the Institute of Industrial Sciences and Process Management and director of the Research Center for Information System Design at the University of Kassel.