Seeing, Hearing, Feeling: Designing Multimodal Alerts for Critical Drone Scenarios

Nina Knieriemen
Saarland University & DFKI
Saarland Informatics Campus
Saarbrücken, Germany
nina.knieriemen@uni-saarland.de

Anke Hirsch
Saarland University & DFKI
Saarland Informatics Campus
Saarbrücken, Germany
anke.hirsch@uni-saarland.de

Muhammad Moiz Sakha
Saarland University & DFKI
Saarland Informatics Campus
Saarbrücken, Germany
moiz.sakha@dfki.de

Florian Daiber
DFKI
Saarland Informatics Campus
Saarbrücken, Germany
florian.daiber@dfki.de

Hannah Kolb
Saarland University
Saarland Informatics Campus
Saarbrücken, Germany
hannah.kolb@dfki.de

Simone Hüning
Saarland University
Saarland Informatics Campus
Saarbrücken, Germany
simone.huening@dfki.de

Frederik Wiehr
Saarland University & DFKI
Saarland Informatics Campus
Saarbrücken, Germany
frederik.wiehr@dfki.de

Antonio Krüger
DFKI & Saarland University
Saarland Informatics Campus
Saarbrücken, Germany
antonio.krueger@dfki.de

Abstract

Autonomous drones are increasingly used across various domains, yet critical situations can arise, and little research exists on how users prefer to be alerted during these events. In multi-drone control scenarios, where human-machine interfaces are used to monitor multiple drones simultaneously, alerting preferences are critical for ensuring situational awareness and timely decision-making. This paper explores multimodal alert design preferences in a user-centered approach. In an online survey, drone pilots identified critical scenarios, with collision risks, signal loss, and hardware problems being the most prevalent challenges. The subsequent study examined notification preferences for multi-drone control interfaces. Participants designed alerts for critical scenarios that were created based on the findings from the first survey. Using a printed control room interface with drone feeds and a map view, participants created multimodal alerts combining visual cues (e.g., frames, text), auditory signals (e.g., beeps), and, less frequently, tactile notifications (vibrations). This work bridges real-world drone operation challenges with user-centered multimodal interface design for autonomous systems.

CCS Concepts

• **Human-centered computing → Empirical studies in HCI**.

ICMI '25, Canberra, ACT, Australia

ACM ISBN 979-8-4007-1499-3/25/10
https://doi.org/10.1145/3716553.3750784

Keywords

autonomous drones; multimodal interaction; UAV; drone handover; criticality; human-computer interaction; user study; elicitation study; qualitative analysis

ACM Reference Format:
Nina Knieriemen, Anke Hirsch, Muhammad Moiz Sakha, Florian Daiber, Hannah Kolb, Simone Hüning, Frederik Wiehr, and Antonio Krüger. 2025. Seeing, Hearing, Feeling: Designing Multimodal Alerts for Critical Drone Scenarios. In *Proceedings of the 27th International Conference on Multimodal Interaction (ICMI '25), October 13–17, 2025, Canberra, ACT, Australia.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3716553.3750784

1 Introduction

Drones are increasingly used in many fields, such as agriculture [35], transportation [21], search and rescue [17], or for private purposes [34]. Companies, such as Amazon (Prime Air[1]) and Wing[2] already offer delivery services using drones. As outlined in the Artificial Intelligence Act of the EU[3], the design of high-risk AI systems has to ensure effective human oversight. Drone technology is advancing in two key directions: developing sophisticated autonomous flight planning systems [37, 41], and moving from single-drone operations to multi-drone systems [2, 5, 8, 40]. Overseeing multiple unmanned aerial vehicles (UAVs) at the same time poses challenges in maintaining situational awareness and managing critical events. In such multi-drone control rooms, an effective interaction design of human-machine interfaces (HMIs) is crucial to alert operators in critical situations properly. When poorly designed or excessive, alerts can overwhelm users, possibly leading to delayed responses in time-critical scenarios [4]. Well-designed alerts are necessary to enhance the operators' ability to recognize,

[1] https://www.aboutamazon.com/news/operations/mk30-drone-amazon-delivery-packages, last accessed on August 21, 2025
[2] https://wing.com/, last accessed on August 21, 2025
[3] https://artificialintelligenceact.eu/article/14/, last accessed on August 21, 2025

prioritize, and respond timely to critical events. Despite the increasing prevalence of autonomous drones, research on user-centered designs of alerting mechanisms for multi-drone interfaces remains sparse. In this paper, a user-centered approach is used to explore drone operators' preferred alert designs using real world critical scenarios identified by drone pilots. We provide actionable insights for the development of multimodal HMIs that support safe and efficient multi-UAV operations.

Our work expands upon human-computer interaction (HCI) findings showing the benefits of using a combination of modalities for alerts when an operator needs to take control of a semi-autonomous drone during a critical situation. Gundappa et al. [15] found that combining auditory alerts with visual icons and spoken messages helped operators to be more accurate and better understand critical situations. We focus on interface design for monitoring of several semi-autonomous drones in a control room scenario. In particular, our contributions include the following:

- We identified the most commonly occurring critical scenarios during drone flight by conducting an online user survey with 55 drone pilots (see Section 4);
- We explored which combinations of alerts (visual, auditory, or tactile) work best for control room interfaces monitoring multiple drones in an elicitation lab study (see Section 5).

We worked directly with experienced drone pilots throughout this process.

2 Background

Effective interfaces for monitoring autonomous systems are crucial as operators need to maintain awareness and intervene in critical situations. Prior work has examined both the nature of these critical situations and how interfaces can alert operators when intervention is needed. We review critical scenarios in autonomous systems in both automotive and drone contexts. We then discuss research on multimodal alert interfaces designed to support operator monitoring and intervention in autonomous systems.

2.1 Critical Scenarios in Autonomous Systems

Critical scenarios in autonomous systems have been studied mainly in automotive contexts. Researchers have defined critical scenarios in several ways: any scenario in which a collision occurs [23, 24], or near-collision situations [16], situations containing dangerous behaviors such as pedestrian encounters at high speeds [1], and cases where predefined critical states are reached, such as unsafe proximity to other agents [11]. Researchers have also developed automated methods for the generation of critical scenarios such as DeepCollision, which uses reinforcement learning to discover environmental configurations that could lead to crashes [29], though concerns persist regarding scenario realism. Such automated methods for discovering critical scenarios have not yet been developed for autonomous drone operations.

Unlike autonomous vehicles that operate on constrained road networks, autonomous drone operations face a wider range of safety-critical situations due to navigation in complex airspace. While automotive critical scenarios often focus on immediate physical interactions (e.g. pedestrian collisions), drone operations also involve evolving situations that could lead to system failures or loss of control. For example, unexpected battery drain during flight could develop into a critical scenario if it threatens the drone's ability to return safely, as drones cannot simply stop in place like ground vehicles. A recent blog post by a drone insurance company on insurance claims highlights the most frequent critical scenarios that involve primarily damage to the drone itself and focuses on cases where the drone needs to be replaced[4]. The most common claims include pilot error, loss of data link, accidental damage, bird strike, and accidental loss. Our study addresses a wider range of critical scenarios beyond just physical drone damage. To identify realistic critical scenarios that can occur in semi-autonomous drone operations, we conducted an online user survey with experienced drone operators.

2.2 Multimodal Alert Interfaces in Autonomous Systems

Multimodal alert interfaces in monitoring systems can present information through different types of sensory feedback (such as visual, auditory, or tactile) to alert operators about critical situations that require their attention or intervention. Given the range of critical situations in autonomous systems, effective alert design within monitoring interfaces is crucial for operator intervention as interface design significantly affects the operator's situation awareness [31]. Research on alert interfaces in autonomous systems has mainly focused on autonomous vehicles [19, 20, 22, 26, 30, 32, 45–48]. A key focus has been on takeover requests (TORs)–alerts that prompt operators to resume control of an autonomous system during critical situations. Researchers have explored ways to improve these alerts through different presentation modalities, including visual cues [32], auditory alerts [22], tactile feedback [47], and olfactory signals [30]. Research showed that users preferred multimodal warnings [45], and that takeover requests combining visual, auditory, and/or tactile modalities led to better driver performance than single-modal alerts [19, 20]. This improved performance may be due to the redundancy provided by multiple modalities working in a complementary way [26, 48]. Visual and auditory combination is the most studied type of bimodal takeover requests in HMIs [9, 27, 32], likely because drivers are already accustomed to interacting with vehicles through these interfaces. Research shows that timely audio-visual explanations of upcoming maneuvers improve response times and obstacle avoidance [6].

In the drone context, Gundappa et al. [15] found similar benefits of multimodal alerts while flying a semi-autonomous drone. Participants had no prior drone flying experience and were instructed to make decisions in critical situations while controlling a single semi-autonomous drone. The study examined a language-based takeover request system that combined auditory signals, visual icons, and spoken messages. As in automotive studies, the authors found that this multimodal approach led to better performance: Participants showed higher accuracy and better recognition of critical situations compared to conditions using spoken messages alone or simple auditory signals.

The unique challenges of monitoring multiple semi-autonomous drones in complex environments necessitate specific research on

[4]https://www.coverdrone.com/eu/an-insight-into-coverdrones-2024-claims/, last accessed on August 21, 2025

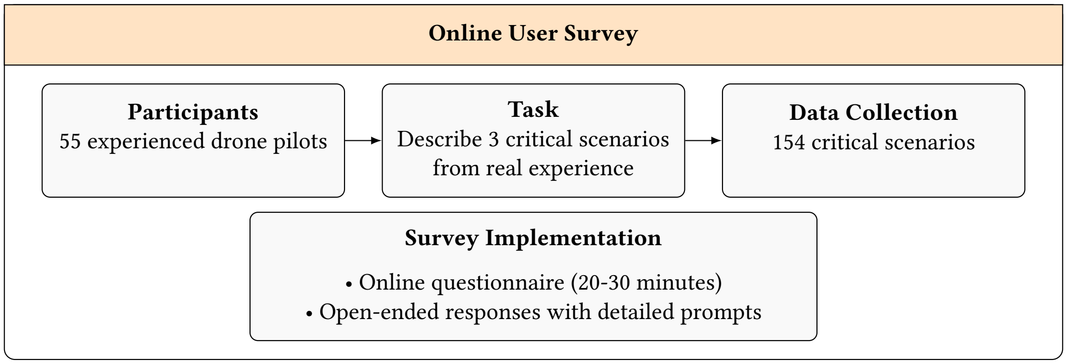

Figure 1: Survey design for eliciting critical scenarios in drone operations. Experienced drone pilots (N = 55) were asked to describe realistic critical situations, resulting in 154 scenario descriptions.

operator alert preferences during critical situations. Therefore, we conducted an elicitation study to understand how operators would prefer to be alerted and their preferences for alerting modalities and visual interface design in control room environments.

2.3 Alerting Systems in Aviation

The design of alerting systems in aviation provides important insights for drone alert interfaces. Research by Lilliboe et al. [28] showed that pilots respond faster to taped voice warnings than to equivalent visual warnings, highlighting the importance of modality in alert design.

The effectiveness of audio alerts depends greatly on how they are presented. Studies have shown that pilots prefer audio alerts that can be canceled, as continuous alerts are often perceived as excessive and disruptive to communication and problem-solving tasks [42]. Ulfvengren et al. [42] also found that auditory icons and animal sounds are easier to learn and that sounds with higher associability to their functions result in faster pilot responses.

Another challenge in aviation alert systems is the proliferation and poor standardization of alerts, which can lead pilots to view them as more of a nuisance than a help [4]. To address this, Patterson et al. [33] proposed a multi-level priority framework for civil aircraft warning systems. This framework categorizes alerts into three urgency levels: immediate-action warnings with the highest priority using a limited set of compatible sounds, lower-priority alerts using *attensons* (attention-getting sounds), and voice alerts integrated into the overall system.

For visual alerts, systems that provide clear instructions on what actions to take and indicate which systems remain operational have been found to better support pilots in high-stress scenarios [38]. These displays present specific actions and provide feedback by removing the instruction once the action is completed.

3 Research Process

Our research process comprised two studies to understand alert design preferences for critical situations in multi-UAV control rooms. To present participants with valid critical situations, we asked drone pilots in a first online survey to provide detailed descriptions of critical drone flight scenarios. In the second user study, we focused on the scenarios most frequently mentioned in the first survey and invited drone pilots to choose their preferred modalities to signal such critical situations during flight. The studies were approved by our institution's Ethical Review Board.

4 Online User Survey

In our online user survey, we asked people who are experienced in flying drones to provide examples of critical scenarios that can occur or have actually occurred while flying a drone. Figure 1 shows our survey design to identify critical scenarios in autonomous drone operations.

4.1 Method

4.1.1 Sample. We recruited 59 participants through mailing lists and by reaching out to companies, civil protection agencies, and volunteer organizations that use drones. Two participants were excluded due to invalid responses and two further participants were excluded because they answered "no" to the question "Do you have experience in flying a drone?". Of the remaining 55 participants, 53 identified as male and 2 as female (age: M = 45, SD = 14.16, range: 18-75). The online questionnaire took 20-30 minutes to complete and was available in both English and German. Participants could enter a raffle to win 20€, with email addresses stored separately to maintain anonymity.

Fourty-eight participants indicated that they have a formal permit to fly a drone in their country. Two participants reported that they fly drones seldomly. Twenty-three fly drones occasionally,

Table 1: Distribution of drone applications reported by participants.

Application Type	Frequency
Agricultural	41
Media/PR	21
Safety/Surveillance	9
Survey/Photogrammetry	9
Recreation	9
Documentation	8
Education/Research	5

Note: Multiple responses per participant were possible

24 participants often fly drones and six participants reportedly fly drones almost every day. Six participants have flown multiple drones at the same time.

The participants reported multiple applications for their drone operations, as shown in Table 1. Agricultural work was the most frequently mentioned, with 41 participants reporting agricultural applications, including locating roe fawns in fields scheduled for cropping (36) and other agricultural tasks such as field spraying or disease prevention (5). Professional applications were also common, covering media production and public relations (21), surveillance and safety monitoring (9), surveying and photogrammetry (9), and infrastructure documentation (8). Five participants reported educational and research applications. Recreational uses included general hobby activities (8) and drone racing (1).

4.1.2 Procedure & Task. The survey was conducted with a LimeSurvey[5] questionnaire. Participants had to read and agree to data privacy terms and regulations. After demographic questions and questions about their drone usage, the instructions for the task followed:

"Imagine your drone flies autonomously and you monitor it. At some point, the drone encounters a problematic situation. Try to think of a realistic and likely situation that could have or actually happened to you. Please describe the situation as detailed as possible. Where does it take place? What are the conditions? What is the environment? What led to the situation?"

Each participant received this instruction three times and could describe three distinct critical situations in open text fields. Some participants reported fewer than three situations, resulting in a total of 154 critical situations.

4.2 Results

4.2.1 Intercoder Reliability. For the analysis of the open text field results, two coders agreed on codes, such as *signal loss* or *collision with a tree*, that they created inductively based on the collected responses [36]. Subsequently, each coder independently assigned the responses to the corresponding codes. The intercoder reliability was calculated using *Cohen's Kappa* which measures the agreement between two independent coders and accounts for the agreement occuring by chance, with values closer to 1 indicating stronger agreement [10]. According to Landis and Koch [25], a Kappa value between 0.61 and 0.80 indicates *substantial* strength of agreement while values between 0.81 and 1.00 represent the highest category described as *almost perfect*. Thus, the intercoder reliability in our study can be considered *substantial* exceeding the threshold of 0.61 with a *Cohen's Kappa* of 0.73 and is therefore sufficiently high to interpret the results [25]. Differences in code assignments were discussed, and the coders agreed on a common assignment. A detailed codebook with all created codes and frequencies can be found in the supplementary material[6].

4.2.2 Code Frequencies. In 37 of the 154 critical situations, the participants reported an environmental context that can be categorized as *natural*, e.g. over a field, forest, mountains or a lake. In 12 situations, an urban context was mentioned, such as construction zones or cities. In 24 situations, bad or impeding weather, such as strong winds, rain, or fog, were mentioned.

Table 2: Most frequently mentioned critical situations.

Situation Type	Frequency
Collision	56
Pilot error	36
Signal loss	24
Hardware issues	19
Battery issues	17
Software issues	15
Connectivity issues	14
Flyaway*	10

*Flyaway refers to drone becoming suddenly uncontrollable.

A summary of the most frequently mentioned critical situations is shown in Table 2. Collision was mentioned the most (in 56 situations). It included collisions with trees, birds, helicopters, and power lines. In 36 situations, a pilot error, such as ignoring warnings or errors in mission planning, was described. Signal loss, for example, between the controller and the drone or a bad GPS signal, was mentioned 24 times. In 19 situations, hardware problems were reported, such as damaged electronics or propellers. Battery problems and problems related to the return of the drone (e.g. Return-to-Home cannot be activated) were reported in 17 situations each. Software problems, such as a software crash were described in 15 situations and not receiving an answer from the drone was mentioned 14 times. In 10 situations, a flyaway (the drone suddenly becomes unable to maneuver and flies away for no apparent reason) was reported as a problematic situation. Other problems were described that were mentioned only once. For example, the safety functions would not work or the drone would not hold the position.

5 Elicitation Study

In the second study, we used an elicitation approach [3, 44] showing drone pilots the scenarios from the online survey and asking them about their preferred modalities and interface design for alerts, as shown in Figure 2.

5.1 Method

5.1.1 Sample. We recruited eight participants (all male; age: M = 36.88, SD = 10.10, range: 23-53) through mailing lists and by reaching out to companies, civil protection agencies, and volunteer organizations using drones. We also distributed flyers on campus. All participants had prior drone flying experience and were fluent in German, as the study was conducted in German to maintain consistency in data collection and analysis. In-person sessions lasted approximately 45 minutes and participants were compensated for their time. All participants indicated that they use some DJI drone model[7]. Four of the participants fly a drone seldomly, one occasionally, one often, and two participants fly a drone almost every day.

[5] https://www.limesurvey.org, last accessed on August 21, 2025
[6] See supplementary material, file: codebooks_studies, sheet: survey
[7] https://www.dji.com/, last accessed on August 21, 2025

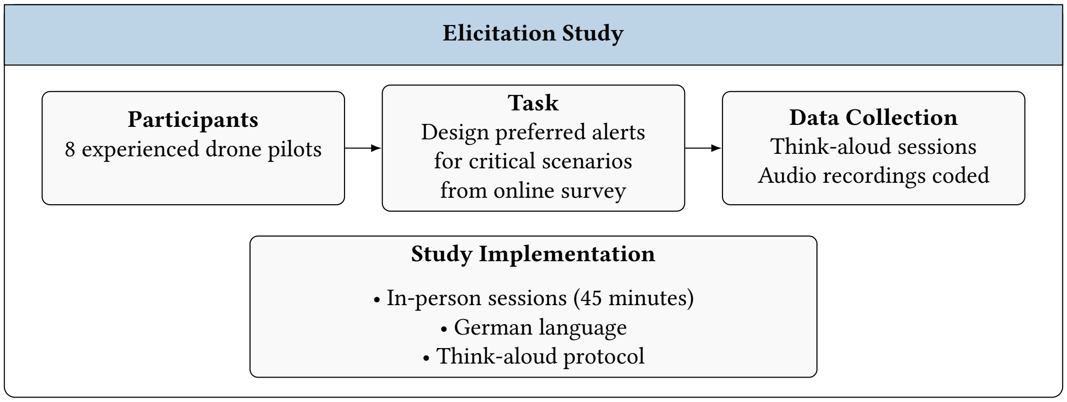

Figure 2: Study design for eliciting preferred alert modalities. Experienced drone pilots (N = 8) designed multimodal alerts for critical scenarios identified in online survey.

5.1.2 Procedure&Task. After reading and agreeing to data privacy terms and regulations, the participants completed a short demographic questionnaire on LimeSurvey[8]. Subsequently, the experimenter described the study setup and led the participant through the study with standardized instructions. Participants were asked to imagine being a drone operator in a control room supervising multiple drones at once that deliver packages or look for fawns in fields. They were shown a paper prototype in A3 format of the splitscreen user interface (see Figure 3). On the left side of the user interface, there is a 3x3 grid displaying the first person view of each drone, and on the right side there is a map view showing the location of each drone.

In four of these drones, one after the other, a critical situation arises in which the participants have to decide how they want to be informed about it. We used the findings of our first survey to create these critical situations based on the most mentioned problem statements in Section 4.2. The most mentioned problem statements were paired with the most mentioned environments and weather conditions. The resulting critical situations were as follows (translated from German):

1. Tree Collision: Your drone flies autonomously over a field. It is slightly cloudy and the visibility is good. The drone is flying towards a tree. There is a risk that the drone will collide with the tree.

2. Signal Loss: Your drone is flying autonomously over hilly terrain. There is a strong crosswind. There is a risk of the drone drifting beyond the controllable range of the controls.

3. Bird Collision: Your drone flies autonomously over a meadow. It is slightly cloudy and the visibility is good. A bird of prey appears in front of the drone. There is a risk that the drone will collide with the bird of prey.

4. Insufficient Battery: Your drone flies autonomously over a forest. The temperature is between -10 and 0°C. The drone's battery suddenly loses power. There is a risk that the remaining battery capacity will not be sufficient to reach the ground station.

We have not included pilot error as drones fly autonomously in this study and pilot errors only lead to further errors, which are then relevant. For example, a participant described the scenario in which the terrain model was not taken into account when planning the flight, and the flight path was then closer to a tree than expected. In this case, the system in our study informs the drone pilot solely about the tree that is in the way and not about the pilot error that happened beforehand and is not of interest to us. The order in which the four situations were shown to the participants was pseudo-randomized with a Latin square. Before starting with a training phase, participants were introduced to the design space of different modalities and possibilities that they had to be alerted by the system.

For the visual modality, we prepared a desk with blank white paper, different (colored) pens, and post-its of different shapes, sizes,

[8]https://www.limesurvey.org, last accessed on August 21, 2025

Figure 3: Split screen paper prototype that participants interacted with in the second user study.

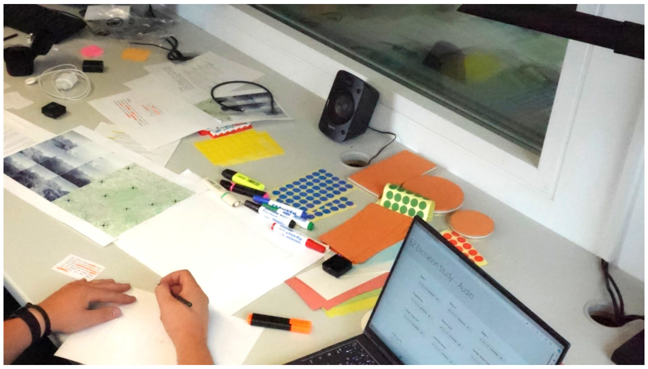

Figure 4: A participant during the elicitation study drawing something while wearing a smartwatch. On the table, we can see a printed drone interface and a laptop with the alert soundboard.

and colors, as you can see in Figure 4. For the auditory modality, participants were presented with nine different alerting sounds and a text-to-speech (TTS) interface on a laptop. The presented sounds can be categorized into immediate alert sounds[9], and attention getting sounds[10], based on the recommendations from Patterson et al. [33]. The sounds were displayed in a 3x3 grid with play buttons so that participants could play each sound as many times as they wanted. On the TTS interface, they could enter the text, give it a title, and select either a female or a male voice from a drop-down list to read the text aloud. For the tactile modality, participants could select from nine different vibration patterns on an Apple watch SE (GPS), second generation, in the App Haptics - Test Haptic Feedback [12]. For each modality, participants were further encouraged to describe what they wanted to hear/see/feel if the options presented did not match their wishes. The design space was therefore not limited by what we presented to them or how well they were able to draw.

After all modalities had been explored, the training phase began with a situation that was the same for all participants, as it was not taken into account for the evaluation: *Your drone flies autonomously over a field. There is early morning fog. Moisture is entering the drone's electronics and threatens to damage them. There is a risk that the drone will crash.* After highlighting the most important aspects of the situation text, participants used the modalities while explaining in a think-aloud process how they would like to be alerted about the specific situation. Subsequently, they repeated this procedure in all four situations. The participants were audio recorded and filmed from behind to capture what they were drawing. Since each of the eight test subjects selected modalities for four situations, we obtained 32 data points.

In Figure 4, you can see that a participant is drawing something while wearing our smartwatch for tactile feedback. The material for the visuals is laid out on the table in front of him together with the printed drone interface and a situation description that has already been highlighted. On a laptop, the soundboard can be seen with the nine alert sounds that we prepared.

[9]e.g., https://freesound.org/people/austin1234575/sounds/213795/, last accessed on August 21, 2025
[10]e.g., https://freesound.org/people/anthonychartier2020/sounds/560188/, last accessed on August 21, 2025
For better readability, the remaining sound links can be found in the alert_sounds.pdf in the supplementary material

Table 3: Distribution of alert modalities in think aloud sessions (N=32 situations).

Modality Type	Design Element	Frequency
Visual	Colors	26
	Icons	21
	Frame	18
	Text Info	16
Auditory	Beep	23
	Text-to-Speech	19
Tactile	Vibration	22

Note: Multiple design elements could be selected for each situation.

5.2 Results

The audios of the think aloud session were transcribed and coded using MAXQDA 24 [43]. Two coders deductively created codes resulting from the modalities available in the study, such as *visual*, *auditory*, and *tactile*. Some more codes were added inductively after reading the first interview [36]. The coders independently assigned the participants' answers to each critical situation to the appropriate codes [25]. With a *Cohen's Kappa* of 0.62, the inter-coder reliability can be considered substantial and is sufficiently high to interpret the results [25]. Differences in code assignments were discussed, and the coders agreed on a common assignment. A detailed codebook with all created codes and frequencies can be found in the supplementary material[11]. Table 3 summarizes the distribution of the alert modalities in the think aloud session, which we examine in detail in the following sections.

All participant quotes presented in the following sections have been translated from German to English.

5.2.1 Visual. The participants created something visual in all 32 situations. Table 4 presents the visual feedback elements that the participants created, with frequency and implementation examples. In 26 situations, participants chose colors to signal something. For example, one participant said: "If there is a hint such as the bird collision, that would be yellow. But if there is a critical problem where landing is required, then I would color it red". In 21 situations, they described one or more icons that display, for example, a bird in critical situation 3, or a battery symbol with an exclamation mark in critical situation 4. In 18 situations, they wanted a frame, either around the drone on the map (mentioned in 2 situations), or around the first person view (11), or both (4). One participant described: "And if a red frame appears around drone five, then I also know which drone is affected". In 16 situations, they wanted to read an info text on the screen, such as "Collision risk" in critical situation 1. An example statement for an overlay with info/icons (3) reads: "That would now be something general, that these are layers that can be switched on and off. And if a warning occurs, perhaps a configuration is selected for this warning category that fits. For example, in the case of wind warning, the wind map fits.". Zooming in or out has also been mentioned (7) as a visual element,

[11]See supplementary material, file: codebooks_studies, sheet: elicitation study

Table 4: Visual feedback elements mentioned by the participants (N=32 situations).

Visual Element*	Frequency	Example Implementations
Colors	26	Color-coded alerts
Icons	21	Bird symbol Battery alert
Frames	18	Frame around first-person view Frame around drone on map
Text Info	16	Collision risk!
Zooming	7	Automatic zoom in/out
Overlay	3	Icon overlay on main display
Flight Path	1	Visualization of drone's trajectory

Note: Multiple visual elements could be selected for each situation.
*All icon images used in this table are taken from Flaticon.com.

for example: "So, I would make the camera image larger somehow to raise alarm" (critical situation 2). Another visual element described is the flight path of the drone: "I would like to have a detailed map of where the drone is and its planned route. [...] Also the planned route and not just the history".

5.2.2 Auditory. In all 32 situations, participants chose the auditory modality. In 23 scenarios, a beep signal was described, for example: "[It] would be good if there was a signal, both physical and auditory. I think a simple beep would suffice here" (critical situation 2). In 8 of those 23 situations, participants chose sounds from the soundboard we showed them.

Participants requested text-to-speech information in 19 situations and created specific text content for 14 situations. For example, one participant created the TTS info "Bird warning! Danger of drone colliding with a bird" for critical situation 3. Another participant typed in: "Battery warning! Power drop leads to insufficient capacity to reach the destination" for critical situation 4. Of those 14 TTS texts, participants chose 4 in a female voice and 7 in a male voice. In 3 situations, they just created a text without actually choosing a voice and playing it. No specific TTS-info was created in the other 5 situations. The participants solely explained that they want to hear spoken information, for example: "A warning is definitely not a bad idea. This means [I want] both the audio signal and the spoken text of the collision warning" (critical situation 3).

5.2.3 Tactile. In 22 of the 32 situations, a vibration was described. In 8 situations, a vibration pattern was selected from the presented smart watch app. In the other 14 scenarios, the participants mentioned that they want to be alerted by a vibration without providing specific details. For example, one participant said for critical situation 2: "So we assume that we want to prevent [the drone] from getting out of reach. Then of course audio and text again. And similar to the first scenario, that you also get vibration feedback after a certain time, let's just say ten seconds. But especially if you do not get the first two audio and text cues, you should also get vibration alerts". Another participant described a red frame around the drone on the map and around the first-person view whenever the criticality was high and said: "In general, if you have a red frame, there is always a vibration with it".

5.2.4 General. All participants indicated that they would like to have a change in the cues adapted to the situation's criticality. This was reflected, for example, in a longer, more frequent, or louder beep signal when the situation was more critical. It was also expressed in an added or more intense/longer vibration or a change in the chosen colors from green to orange to red. For example, one participant said about critical situation 4: "I actually probably see the warning level as lower here, haptics and beep, because although this is a serious problem, it is less urgent. [...] This is a slightly lower level of urgency because the battery temperature does not change suddenly. [...] That means [I choose] the slightly less prominent haptic, but hopefully still one that is well noticeable. One of the easily audible but less critical warning beeps". Two people requested fixed warnings for specific cases as in aviation. For example, the warning for increased wind or low battery should always sound, look, and feel exactly the same so that they immediately know what to do.

6 Discussion

In our research, we gained insights about the design of HMIs in a multi-drone control room scenario. Although earlier studies showed that multiple modalities lead to better user performance [15, 20], they have not explicitly included user groups' preferences in a user-centered elicitation approach. Our findings suggest that drone pilots want to be alerted to a critical situation through multiple modalities. The visual and auditory modalities were the most popular in our study, both of which, according to our findings, should always occur in an alert, closely followed by the tactile modality. The most mentioned visual notification cues were colors, icons, frames, and an info text, which were to be paired with auditory signals such as a beep and subsequent TTS info. Based on our findings, it can be assumed that vibration should also be included in most cases, especially if the situation described is more critical.

6.0.1 Use of Multimodal alerts. Our analysis revealed the consistent use of multiple modalities across critical scenarios. Operators consistently created multimodal alert designs. Visual and auditory elements were included in all designs, with tactile feedback incorporated in most cases. This preference for combining multiple modalities aligns with findings from autonomous vehicle research, where different modalities have been shown to work in complementary ways [26, 48]. Thus, the results imply that it is recommendable to combine multiple modalities, particularly visual and auditory alerts. The additional integration of tactile feedback is also recommended, especially in situations with higher criticality.

6.0.2 Adapt Alerts to Criticality. All participants indicated a need for alerts that adapt to the criticality of the situation. This was reflected in their designs through progressive intensification across modalities: increasing audio frequency or volume, more intense vibration patterns, and color progression from green to orange to red. For auditory alerts, this finding aligns with research on automotive alert interfaces, where signal characteristics can convey different levels of urgency. The perceived urgency increases with faster pulse rates [13, 39] and higher frequencies [14], while it decreases with slower pulse rates and lower frequencies. This result indicates that it is advisable to adjust the intensity of the alert to the criticality of the situation.

6.0.3 Structure Alerts from Immediate Signals to Detailed Information. Within each modality, the operators structured alerts hierarchically to convey different levels of information. In the visual domain, most of the user designs combined immediate signaling elements (colors, frames) with more detailed information layers. For example, designs commonly progressed from color indicators to specific icons and supplementary text. This layering was particularly evident in the critical battery scenario, where operators sequenced red warning colors with battery icons and numerical indicators. The audio designs followed a similar hierarchical pattern, with the majority of the participants using alert tones for immediate attention followed by voice messages for detailed information. These findings suggest that it makes sense to use immediate cues, such as beeps, colors, and frames to capture the user's attention, followed by more detailed information, such as TTS, or textual information.

6.0.4 Incorporate Standardized Alerts. Additionally, two operators suggested incorporating standardized alerts for specific critical scenarios, similar to aviation protocols, where consistent multimodal patterns would signal specific conditions such as low battery or high winds. This recommendation aligns with historical lessons from aviation, where due to the proliferation of alerts and the lack of standardization, airline pilots began to view alerting systems as a nuisance rather than a help [4]. It is therefore advisable to use consistent multimodal patterns for specific situations to enable fast recognition and reaction of the user.

6.1 Limitations and Future Work

Due to specific participant requirements—including drone operation experience, German language proficiency, and the need to conduct a controlled study in person—our elicitation study had a limited sample size. While similar sample sizes have been used in some drone HCI studies (e.g. [7, 18]), our relatively small sample remains a constraint. The limited availability of drone operators made recruitment challenging; however, the consistency in their design preferences suggests valuable insights despite the sample size. In addition, in the lab study, participants were asked to imagine monitoring drones and interact with paper prototypes. This might lack the dynamic of a real control room environment. In the future, we plan to conduct user studies using a drone simulator to present critical scenarios with the preferred modalities identified in our current study, and assess reaction time, situational awareness, cognitive load, and other measures.

7 Conclusion

Our results provide useful insights into user preferences for notification modalities when monitoring autonomous drones. Designing interfaces that effectively support operators by enhancing their situational awareness while minimizing cognitive load remains a pressing challenge. Our findings support the potential of a multimodal alert system that incorporates visual, auditory, and tactile cues. By integrating drone pilot input into both the identification of critical situations during a drone flight and the design of alerts, we bridge the gap between real-world drone operations and user-centered interface design.

Safe and Responsible Innovation Statement

This research promotes safe and responsible human-drone interaction by focusing on user-centered design of multimodal alerts in drone operations. The alert design is essential to increase situational awareness among users and thus ensure safe control of the drone and minimize the risk of accidents. We acknowledge that improved alerting systems in multi-drone control interfaces could be misused for military purposes. Our design process centered non-combative use cases such as search and rescue and we strongly oppose the use of this research for combative applications. We encourage future adopters to critically examine the context of use and ensure that it is consistent with humanitarian, non-military values. The data of the two user studies were processed anonymously and both studies were approved by our institution's Ethical Review Board.

Acknowledgments

This work was supported by DFG grant 389792660 as part of TRR 248 - CPEC.

References

[1] Raja Ben Abdessalem, Shiva Nejati, Lionel C. Briand, and Thomas Stifter. 2018. Testing Vision-Based Control Systems using Learnable Evolutionary Algorithms. In *Proceedings of the 40th International Conference on Software Engineering* (Gothenburg, Sweden) *(ICSE '18)*. Association for Computing Machinery, New York, NY, USA, 1016–1026. https://doi.org/10.1145/3180155.3180160

[2] Ankit Agrawal, Sophia J. Abraham, Benjamin Burger, Chichi Christine, Luke Fraser, John M. Hoeksema, Sarah Hwang, Elizabeth Travnik, Shreya Kumar, Walter Scheirer, Jane Cleland-Huang, Michael Vierhauser, Ryan Bauer, and Steve Cox. 2020. The Next Generation of Human-Drone Partnerships: Co-Designing an Emergency Response System. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376825

[3] Keith C. Barton. 2015. Elicitation Techniques: Getting People to Talk About Ideas They Don't Usually Talk About. *Theory & Research in Social Education* 43, 2 (2015), 179–205. https://doi.org/10.1080/00933104.2015.1034392

[4] B. L. Berson, D. A. Po-Chedley, G. P. Boucek, D. C. Hanson, and M. F. Leffler. 1981. *Aircraft Alerting Systems Standardization Study. Volume II. Aircraft Alerting System Design Guidelines.* Technical Report FAA/RD-81/38. Boeing Commercial Airplane Co., Seattle, WA. 188 pages. https://ntrl.ntis.gov/NTRL/dashboard/searchResults/titleDetail/ADA106732.xhtml

[5] Oscar Bjurling, Rego Granlund, Jens Alfredson, Mattias Arvola, and Tom Ziemke. 2020. Drone Swarms in Forest Firefighting: A Local Development Case Study of Multi-Level Human-Swarm Interaction. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society* (Tallinn, Estonia) *(NordiCHI '20)*. Association for Computing Machinery, New York, NY, USA, Article 93, 7 pages. https://doi.org/10.1145/3419249.3421239

[6] Shadan Sadeghian Borojeni, Lars Weber, Wilko Heuten, and Susanne Boll. 2018. From Reading to Driving: Priming Mobile Users for Take-over Situations in Highly Automated Driving. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Barcelona, Spain) *(MobileHCI '18)*. Association for Computing Machinery, New York, NY, USA, Article 14, 12 pages. https://doi.org/10.1145/3229434.3229464

[7] Jessica R. Cauchard, Alex Tamkin, Cheng Yao Wang, Luke Vink, Michelle Park, Tommy Fang, and James A. Landay. 2020. Drone.io: A Gestural and Visual Interface for Human-Drone Interaction. In *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction* (Daegu, Republic of Korea) *(HRI '19)*. IEEE Press, New York, NY, USA, 153–162. https://doi.org/10.1109/HRI.2019.8673011

[8] Serge Chaumette and Frédéric Guinand. 2017. Control of a Remote Swarm of Drones/Robots Through a Local (Possibly Model) Swarm: Qualitative and Quantitative Issues. In *Proceedings of the 14th ACM Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, & Ubiquitous Networks* (Miami, Florida,

USA) *(PE-WASUN '17)*. Association for Computing Machinery, New York, NY, USA, 41–45. https://doi.org/10.1145/3134829.3134840

[9] Kuan-Ting Chen, Huei-Yen Winnie Chen, and Ann Bisantz. 2023. Adding Visual Contextual Information to Continuous Sonification Feedback about Low-Reliability Situations in Conditionally Automated Driving: A Driving Simulator Study. *Transportation Research Part F: Traffic Psychology and Behaviour* 94 (2023), 25–41. https://doi.org/10.1016/j.trf.2023.01.017

[10] Jacob Cohen. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* 20, 1 (1960), 37–46. https://doi.org/10.1177/001316446002000104

[11] Anthony Corso, Peter Du, Katherine Driggs-Campbell, and Mykel J. Kochenderfer. 2019. Adaptive Stress Testing with Reward Augmentation for Autonomous Vehicle Validation. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)* (Auckland, New Zealand). IEEE Press, Piscataway, NJ, 163–168. https://doi.org/10.1109/ITSC.2019.8917242

[12] Adam Foot. 2019. *Haptics - Test Haptic Feedback*. Durham, United Kingdom. https://apps.apple.com/de/app/haptics-test-haptic-feedback/id1474606532 Mobile application.

[13] Christian Gonzalez, Bridget A. Lewis, Daniel M. Roberts, Stephanie M. Pratt, and Carryl L. Baldwin. 2012. Perceived Urgency and Annoyance of Auditory Alerts in a Driving Context. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 56, 1 (2012), 1684–1687. https://doi.org/10.1177/1071181312561337

[14] A Guillaume, C Drake, M Rivenez, L Pellieux, and V Chastres. 2002. Perception of Urgency and Alarm Design. In *Proceedings of the 8th International Conference on Auditory Display*. Advanced Telecommunications Research Institute Kyoto, 357–361.

[15] Ashwini Gundappa, Emilia Ellsiepen, Lukas Schmitz, Frederik Wiehr, and Vera Demberg. 2024. The Design of Informative Take-Over Requests for Semi-Autonomous Cyber-Physical Systems: Combining Spoken Language and Visual Icons in a Drone-Controller Setting. arXiv:2409.08253

[16] Florian Hauer, Alexander Pretschner, and Bernd Holzmüller. 2019. Fitness Functions for Testing Automated and Autonomous Driving Systems. In *Computer Safety, Reliability, and Security: 38th International Conference, SAFECOMP 2019, Turku, Finland, September 11–13, 2019, Proceedings* (Turku, Finland). Springer-Verlag, Berlin, Heidelberg, 69–84. https://doi.org/10.1007/978-3-030-26601-1_5

[17] Yao-Hua Ho and Yu-Jung Tsai. 2022. Open Collaborative Platform for Multi-Drones to Support Search and Rescue Operations. *Drones* 6, 5 (2022). https://doi.org/10.3390/drones6050132

[18] Maria-Theresa Oanh Hoang, Niels van Berkel, Mikael B. Skov, and Timothy R. Merritt. 2023. Challenges and Requirements in Multi-Drone Interfaces. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 65, 9 pages. https://doi.org/10.1145/3544549.3585673

[19] Gaojian Huang and Brandon J. Pitts. 2022. The Effects of Age and Physical Exercise on Multimodal Signal Responses: Implications for Semi-Autonomous Vehicle Takeover Requests. *Applied Ergonomics* 98 (2022), 103595. https://doi.org/10.1016/j.apergo.2021.103595

[20] Gaojian Huang, Clayton Steele, Xinrui Zhang, and Brandon J. Pitts. 2019. Multimodal Cue Combinations: A Possible Approach to Designing In-Vehicle Takeover Requests for Semi-Autonomous Driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 63 (2019), 1739–1743. https://doi.org/10.1177/1071181319631053

[21] Robin Kellermann, Tobias Biehle, and Liliann Fischer. 2020. Drones for Parcel and Passenger Transportation: A Literature Review. *Transportation Research Interdisciplinary Perspectives* 4 (2020). https://doi.org/10.1016/j.trip.2019.100088

[22] Sangjin Ko, Harsh Sanghavi, Yiqi Zhang, and Myounghoon Jeon. 2022. Modeling the Effects of Perceived Intuitiveness and Urgency of Various Auditory Aarnings on Driver Takeover Performance in Automated Vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour* 90 (2022), 70–83. https://doi.org/10.1016/j.trf.2022.08.008

[23] Mark Koren, Saud Alsaif, Ritchie Lee, and Mykel J. Kochenderfer. 2018. Adaptive Stress Testing for Autonomous Vehicles. In *2018 IEEE Intelligent Vehicles Symposium (IV)* (Changshu, Suzhou, China). IEEE Press, 1–7. https://doi.org/10.1109/IVS.2018.8500400

[24] Mark Koren and Mykel J. Kochenderfer. 2019. Efficient Autonomy Validation in Simulation with Adaptive Stress Testing. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. 4178–4183. https://doi.org/10.1109/ITSC.2019.8917403

[25] J. Richard Landis and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics: Journal of the International Biometric Society* 33, 1 (1977), 159–174. https://doi.org/10.2307/2529310

[26] Okkeun Lee and Hyunmin Kang. 2023. Individual Differences in Signal Perception for Takeover Request in Autonomous Driving. *Applied Sciences* 13, 14 (2023). https://doi.org/10.3390/app13148162

[27] Shuo Li, Phil Blythe, Weihong Guo, Anil Namdeo, Simon Edwards, Paul Goodman, and Graeme Hill. 2019. Evaluation of the Effects of Age-Friendly Human-Machine Interfaces on the Driver's Takeover Performance in Highly Automated Vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour* 67 (2019), 78–100. https://doi.org/10.1016/j.trf.2019.10.009

[28] Monard L. Lilleboe. 1963. Evaluation of Astropower, Inc., Auditory Information Display Installed in the VA-3B Airplane. Technical Report ST 31-22R-63.

[29] Chengjie Lu, Yize Shi, Huihui Zhang, Man Zhang, Tiexin Wang, Tao Yue, and Shaukat Ali. 2023. Learning Configurations of Operating Environment of Autonomous Vehicles to Maximize their Collisions. *IEEE Transactions on Software Engineering* 49, 1 (2023), 384–402. https://doi.org/10.1109/TSE.2022.3150788

[30] Manik Mahachandra, Yassierli, and Erdo D. Garnaby. 2015. The Effectiveness of In-Vehicle Peppermint Fragrance to Maintain Car Driver's Alertness. *Procedia Manufacturing* 4 (2015), 471–477. https://doi.org/10.1016/j.promfg.2015.11.064

[31] Jakob Nielsen. 1994. *Usability Engineering*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[32] Yang-Kun Ou, Wan-Xuan Huang, and Chen-Wen Fang. 2021. Effects of Different Takeover Request Interfaces on Takeover Behavior and Performance during Conditionally Automated Driving. *Accident Analysis & Prevention* 162 (2021). https://doi.org/10.1016/j.aap.2021.106425

[33] Roy D. Patterson. 1982. *Guidelines for Auditory Warning Systems on Civil Aircraft.* Technical Report. Civil Aviation Authority, London, England. 110 pages. https://ntrl.ntis.gov/NTRL/dashboard/searchResults/titleDetail/PB83248625.xhtml

[34] Bharat Rao, Ashwin Goutham Gopi, and Romana Maione. 2016. The Societal Impact of Commercial Drones. *Technology in Society* 45 (2016), 83–90. https://doi.org/10.1016/j.techsoc.2016.02.009

[35] Abderahman Rejeb, Alireza Abdollahi, Karim Rejeb, and Horst Treiblmaier. 2022. Drones in Agriculture: A Review and Bibliometric Analysis. *Computers and Electronics in Agriculture* 198 (2022). https://doi.org/10.1016/j.compag.2022.107017

[36] Carol Rivas. 2012. Coding and Analysing Qualitative Data. In *Researching Society and Culture. Third Edition*, Clive Seale (Ed.). SAGE Publications, 367–392.

[37] Davide Di Ruscio, Ivano Malavolta, Patrizio Pelliccione, and Massimo Tivoli. 2016. Automatic Generation of Detailed Flight Plans from High-Level Mission Descriptions. In *Proceedings of the ACM/IEEE 19th International Conference on Model Driven Engineering Languages and Systems* (Saint-malo, France) *(MODELS '16)*. Association for Computing Machinery, New York, NY, USA, 45–55. https://doi.org/10.1145/2976767.2976794

[38] Gideon Singer and Sidney WA Dekker. 2000. Pilot Performance During Multiple Failures: An Empirical Study of Different Warning Systems. *Transportation Human Factors* 2, 1 (2000), 63–76.

[39] Clara Suied, Patrick Susini, and Stephen McAdams. 2008. Evaluating Warning Sound Urgency with Reaction Times. *Journal of Experimental Psychology: Applied* 14, 3 (2008), 201–212. https://doi.org/10.1037/1076-898X.14.3.201

[40] Pietro Tosato, Daniele Facinelli, Maurizio Prada, Luca Gemma, Maurizio Rossi, and Davide Brunelli. 2019. An Autonomous Swarm of Drones for Industrial Gas Sensing Applications. In *2019 IEEE 20th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*. 1–6. https://doi.org/10.1109/WoWMoM.2019.8793043

[41] Yu-Hsuan Tu, Stuart Phinn, Kasper Johansen, Andrew Robson, and Dan Wu. 2020. Optimising Drone Flight Planning for Measuring Horticultural Tree Crop Structure. *ISPRS Journal of Photogrammetry and Remote Sensing* 160 (2020), 83–96. https://doi.org/10.1016/j.isprsjprs.2019.12.006

[42] Pernilla Ulfvengren, Lena Mårtensson, and Gideon Singer. 2001. Auditory and Visual Warnings in Aircraft. *IFAC Proceedings Volumes* 34, 16 (2001), 53–57. https://doi.org/10.1016/S1474-6670(17)41501-1 8th IFAC Symposium on Analysis, Design and Evaluation of Human-Machine Systems (HMS 2001), Kassel, Germany, 18-20 September, 12001.

[43] VERBI Software. 2023. *MAXQDA 2024 (24.6.0)*. VERBI Software, Berlin, Germany. https://www.maxqda.com/

[44] Santiago Villarreal-Narvaez, Jean Vanderdonckt, Radu-Daniel Vatavu, and Jacob O. Wobbrock. 2020. A Systematic Review of Gesture Elicitation Studies: What Can We Learn from 216 Studies?. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS '20)*. Association for Computing Machinery, New York, NY, USA, 855–872. https://doi.org/10.1145/3357236.3395511

[45] Marcel Walch, Kristin Lange, Martin Baumann, and Michael Weber. 2015. Autonomous Driving: Investigating the Feasibility of Car-Driver Handover Assistance. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Nottingham, United Kingdom) *(AutomotiveUI '15)*. Association for Computing Machinery, New York, NY, USA, 11–18. https://doi.org/10.1145/2799250.2799268

[46] Frederik Wiehr, Anke Hirsch, Lukas Schmitz, Nina Knieriemen, Antonio Krüger, Alisa Kovtunova, Stefan Borgwardt, Ernie Chang, Vera Demberg, Marcel Steinmetz, and Jörg Hoffmann. 2021. Why Do I Have to Take Over Control? Evaluating Safe Handovers with Advance Notice and Explanations in HAD. In *Proceedings of the 2021 International Conference on Multimodal Interaction* (Montréal, QC, Canada) *(ICMI '21)*. Association for Computing Machinery, New York, NY, USA, 308–317. https://doi.org/10.1145/3462244.3479884

[47] Sol Hee Yoon, Young Woo Kim, and Yong Gu Ji. 2019. The Effects of Takeover Request Modalities on Highly Automated Car Control Transitions. *Accident Analysis & Prevention* 123 (2019), 150–158. https://doi.org/10.1016/j.aap.2018.11.018

[48] Hanna Yun and Ji Hyun Yang. 2020. Multimodal Warning Design for Take-Over Request in Conditionally Automated Driving. *European Transport Research Review* 12 (2020), 1–11. https://doi.org/10.1186/s12544-020-00427-5