Enhancing Notification Awareness for Online Presenters via a Wrist-Worn Device

René Schäfer  
RWTH Aachen University  
Aachen, Germany  /rschaefer@cs.rwth-aachen.de

Ulyana Lavnikevich  
RWTH Aachen University  
Aachen, Germany  
ulyana.lavnikevich@rwth-aachen.de

Tobias Wagner  
Institute of Media Informatics  
Ulm University, Germany  
tobias.wagner@uni-ulm.de

Jan Borchers  
RWTH Aachen University  
Aachen, Germany  
borchers@cs.rwth-aachen.de

Figure 1: Our wrist-worn prototype supplied presenters with additional visual, haptic, or auditory cues during online presentations to raise awareness of incoming notifications from the audience without disturbing the presenter’s flow. Our study participants had to recognize three different types of Zoom notifications and press a corresponding button while giving a presentation.

ABSTRACT
The practice of giving presentations online has exploded during the Covid pandemic. However, in these settings, presenters often find themselves overlooking questions and feedback, e.g. via chat, from the audience, because the presenter’s screen is dominated by their slides, with other channels becoming less noticeable. This causes frustration among presenters and their audience alike. We investigate the impact of additional visual, auditory, and haptic cues for presenters in online scenarios, using a wrist-worn prototype. For this, we conducted a study where participants gave presentations via the videoconferencing tool Zoom on specific topics while trying to notice and correctly identify incoming notifications. Our findings indicate that supplementary notifications can be helpful in online presentations without inappropriately disturbing the presenter.

CCS CONCEPTS
• Human-centered computing → Haptic devices; User studies.
KEYWORDS
Notification Awareness, Online Presenters, Interaction Modalities, Wrist-Worn, Peripheral Recognition, Audience Feedback

ACM Reference Format:

1 INTRODUCTION
Teaching institutions rely increasingly on online tools and services to deliver course content [16]. However, during the Covid pandemic, switching from in-person to online presentations became inevitable for universities, schools, and other teaching facilities [9, 15]. This put new emphasis on questions of the quality of online presentations and the use of video conferencing tools such as Zoom1 or Teams2.

One key issue that has become apparent is that online presenters tend to focus on their own slides and notes, which occupies their visual channel [8, 20]. This leads to them missing other notifications, such as questions in the text chat, raised hands, and other forms of communication [12]. Cognitively, for the presenter the audience fades into the background. When an important question or comment is missed this way, participants then need to either interrupt the presenter to resolve the issue, or leave their concerns and questions unheard. This can lead to them losing track of the presentation, or even stopping to listen altogether, causing frustration on both sides. However, since giving online presentations mainly occupies the visual channel (e.g., by looking at the slides and presenter notes), other senses may still be available to perceive audience feedback. Even additional visual cues might be viable when placed in the near-peripheral view [7]. Still, any additional cues should keep the distraction during the presentation to a minimum.

Our work, thus, aims at improving the perception of important notifications during online presentations without inappropriately distracting the presenter. To this end, we created a simple wrist-worn prototype that provides additional notification cues using light, sound, or vibrations. Our goal is to investigate whether such additional cues can support presenters without disturbing their presentation flow. For this, we conducted a study where participants gave presentations via Zoom while trying to notice and distinguish incoming notifications. We present insights on missed notifications, and compare how well different feedback modalities are noticed. Our findings indicate that additional feedback cues can be helpful during online presentations and that presenters can distinguish different types of notifications without being overwhelmed. Visual feedback in the near-peripheral view and auditory cues are especially helpful when the presenter is supposed to deduce a specific message, while vibrotactile feedback can be used to make the presenter generally aware of an incoming notification.

2 RELATED WORK
Our work is based on prior research on visual, auditory, and tactile notifications delivered through supplemental devices for presenters engaging in public speaking. Previous investigations have explored the use of notifications conveyed through various modalities, devices, and device placements to enhance presenters’ awareness of their gestural [2, 3] and speaking behavior (e.g., voice pitch, volume, and speaking rate) [1, 3, 18], as well as to support presenters in managing remaining time [1, 17]. Using head-worn glasses [3, 18] or wrist-worn devices [1, 2, 17], feedback was provided through unimodal visual [3, 14, 18] or haptic [2] notifications, or multimodal notifications using both modalities [1]. While head-worn glasses were used primarily to provide visual notifications [3, 18], wrist-worn devices were used for both visual and haptic notifications [1, 14]. For instance, Pohl et al. [14] used LEDs under a smartwatch to provide subtle visual feedback to transmit notifications in more natural and less distracting ways. In the same way, Bubel et al. [1] used an LED display to indicate remaining time and provide haptic feedback as soon as the device detected that the presenter’s anxiety was increasing.

While the above work shows that notifications on supplemental devices can help presenters improve their public speaking skills, it remains unclear how to best provide audience feedback to presenters so that they can react accordingly [6] and without inappropriately distracting them. This becomes especially important when presentations are delivered via video conferencing tools, due to the spatial separation of presenter and audience. In these settings, speaking up and interrupting the presenter is even more challenging than in a face-to-face situation, because of the limitations of current video conferencing technology in handling duplex audio well. As a result, the audience is usually left with only visual feedback channels such as chat messages and emojis, which, however, may overload the presenter’s visual channel [8].

Therefore, we were interested in investigating whether a supplemental wrist-worn device can remedy this problem by utilizing various notification modalities for audience feedback.

3 STUDY
Our study design is based on Warnock et al. [19], but uses different modalities and a remote video conferencing context. We let our participants give actual talks on predetermined topics via the widely used Zoom video conferencing tool.

3.1 Prototype
We built a wrist-worn prototype to test the impact of notifications across different modalities in the peripheral field of view (cf. Fig. 2). It consists of a 12-LED NeoPixel RGB LED ring, a piezo buzzer, a vibration motor, a 3D-printed case, and Velcro. This lets us provide visual, vibrotactile, and auditory feedback. The case has a diameter of 53 mm and is 7 mm thick. An Arduino Uno connected to the prototype triggers all feedback.

3.2 Feedback Variants
Apart from investigating if presenters recognize the arrival of notifications, we also investigated whether they can distinguish between different kinds of notifications. For this, we selected three common
Zoom notifications often used in online presentations: (1) chat messages for asking questions, (2) raising a hand to request permission to speak, and (3) the slow down symbol to tell the presenters that they are too fast to follow.

We studied four different feedback variants in our study: Visual, Tactile, Auditory, and Baseline. The first three involved our prototype; the last included no additional feedback. For each feedback variant, except Baseline, we mapped each of the three Zoom notifications to different stimuli on our prototype.

**Baseline.** In this condition, the prototype was turned off, and participants had to respond solely to the notifications provided by Zoom while presenting. This helped us understand how useful or distracting the additional feedback variants would be, and how much information was missed without further cues.

**Visual.** In this condition, the LED ring glowed in red, yellow, or blue depending on the type of Zoom notification. The actual light pattern we implemented turns on the LEDs on the ring one by one at an interval of 100 ms and keeps them on for 1000 ms when a stimulus is triggered. According to Pohl et al. [14], who investigated subtle notification mechanisms for smartwatches, this illumination mode does not influence reaction time. Of the countless light patterns possible, this combined the two most preferred by participants in [14].

**Tactile.** In this condition, a small vibration motor created three different signals depending on the type of Zoom notification. We followed the design recommendations of Graham-Knight et al. [4] for vibration-based communication, which suggest a combination of three long and similar vibration pulses with a clear distinction between the duration of individual pulses as a suitable representation of a message category. This led to the following vibration patterns for each Zoom notification: (1) 200 ms - 200 ms - 400 ms pulses 300 ms apart, (2) 200 ms - 200 ms - 200 ms pulses 400 ms apart, and (3) 400 ms - 400 ms - 200 ms pulses 200 ms apart.

**Auditory.** Since not all of the tested zoom notifications yield sounds, we used a piezo buzzer that played three different tone sequences depending on the notification type. Following Nault et al. [13], we used a frequency of 800 Hz, and a first tone duration of 100 ms. We added a second and third 500 ms tone at the same frequency for the other notification types, making them comparable to traditional notification alerts. The two-tone sequence used 300 ms pause in between, the three-tone sequence two 100 ms pauses.

### 3.3 Study Setup

We prepared four short presentations on different fairly little-known animals with presenter notes of approximately 1200 words. To minimize any confounding effects, all presentations were in the same domain and structured similarly. The presentations were designed to take at least 5 minutes, to allow us to control the number of notifications appropriately. Our studies took approximately 90 minutes each. Additionally, our setup included our wrist-worn prototype, a laptop with a screen size of 15” that had Zoom installed, an Arduino Uno driving the prototype, and three 3D-printed buttons for the study task as seen in Fig. 1. The screen layout for the study is shown in Fig. 3.

### 3.4 Task and Procedure

Participants were given two tasks. Their primary task was to give presentations on the predetermined topics. The secondary task, following Warnock et al. [19], was to notice and distinguish between incoming notifications and press the corresponding button in front of the laptop. To achieve a scenario and stress level similar to that of an actual online presentation, we told participants that their presentations would be recorded.

We introduced participants to our prototype and to the respective feedback variant it provides. Every participant wore the prototype on their non-dominant hand. Participants familiarized themselves with the current feedback variant. For this, a Zoom notification was triggered using the corresponding feedback of the prototype. Participants then pressed the button for that notification. This process was repeated until participants felt confident and aware of all three possible notification types. Afterwards, they were given 5 minutes to prepare their current presentation. Once they were ready, the presentation started. After 5 minutes, the presentation was stopped to control the mental load. No participant finished a
presentation before this mark. During the presentations, participants received Zoom notifications and had to react by pressing the matching button in front of them. Chat messages were displayed in the chat, while slow-down symbols and raised hands were shown in the top left corner of the field of the respective audience member (Fig. 3). In addition to showing the icon, raised hands move the field of that audience member to the left of the row. Simultaneously with the notification, the prototype provided the matching feedback for the Zoom notification. All buttons were labeled according to the given condition stating the Zoom notification and the stimuli from our prototype. Each participant received six notifications (two for each Zoom notification type) at random points in the presentation with at least 10 s between two stimuli. After the presentation, participants removed the prototype from their wrist and filled out a questionnaire about the current condition. This questionnaire contained the NASA Task Load Index (NASA TLX) [5], the short version of the User Experience Questionnaire (UEQ-S) [10], and fields for comments regarding the given feedback variant. Additionally, participants filled out Likert scales on the ease to interpret notifications, whether they were able to distinguish them, whether they were confident to have noticed all notifications, and whether they felt distracted during the talk. Afterwards, the procedure was repeated for the remaining feedback variants. The prototype was worn during each presentation, including Baseline to make all conditions feel similar. After all presentations, participants answered a final questionnaire where they provided demographic data, ranked all feedback variants, and explained their ranking. Additionally, participants were asked whether they are giving presentations frequently and whether they like doing so.

Overall, our within-subject design led to 4 Presentations × 6 Notifications × 16 Participants, resulting in 384 trials. Notification types were counterbalanced using a Latin Square, and the presentations were randomized.

3.5 Dependent Variables
We measured the reaction time using an Arduino Uno and the correct, incorrect, and missed notifications. Additionally, we collected data on two common questionnaires (NASA TLX and UEQ-S).

4 RESULTS
16 people (age 21 to 30, M = 24.56, SD = 2.98; 8 male, 7 female, 1 non-binary; 1 left-handed) participated in our study. All had an academic background and experience with public speaking online. The conditions using our prototype achieved better results compared to Baseline. In the following, we report results regarding recognition time, successful recognitions, mental load, and the overall ranking.

4.1 Reaction Time
Overall, Auditory had the lowest reaction time to incoming notifications (M = 1.70 s, SD = 0.40 s), followed by Visual (M = 2.03 s, SD = 0.30 s) and Baseline (M = 2.23 s, SD = 0.65 s). Tactile had the highest reaction time (M = 2.85 s, SD = 0.48 s).

4.2 Correct Recognitions
All conditions using the prototype achieved a very high recognition rate (Visual and Auditory 98%, Tactile 96%). Even Baseline achieved more than 84% correct recognitions, but participants missed 14% of all notifications in this condition. Overall, 75% of the participants missed at least one notification without supplemental notifications from our prototype. Participants did not miss any notification in the Tactile and Auditory conditions. With 4%, Tactile had the most recognition mistakes.

4.3 Questionnaires
We evaluated perceived cognitive load during task completion using the NASA TLX [5], see Fig. 4. In addition, we assessed the usability of the wrist-worn device with the UEQ-S [10]. It measures Pragmatic Quality (usefulness and ease of interacting) and Hedonic Quality (joy and emotions triggered by product use). Finally, we assessed how confident our participants felt regarding the overall perception of notifications during the presentations. We summarize our findings below:

Baseline resulted in the highest (most demanding) NASA TLX overall workload score and required a high level of information processing. In particular, it scored highest (worst) in Mental Demand and Effort, and received the second worst score in Performance and Frustration. Surprisingly, although only 14% of notifications were actually missed, half of our participants were still unsure whether they had perceived all incoming signals.

Visual achieved the second best NASA TLX scores across all categories: Mental Demand, Performance, Effort, and Frustration. In the UEQ-S, this modality had the best overall results with a score of 1.57 (Pragmatic Quality: 1.45; Hedonic Quality: 1.69). With this, Visual yielded a positive result (i.e., > 0.8) in Pragmatic and Hedonic Quality. In this condition, about 30% of the respondents overall felt unsure about the number of signals perceived via the LED ring.

Tactile was similar in Mental Demand to Baseline. It had the highest (worst) NASA TLX score in Performance and Frustration. It also had the lowest Pragmatic Quality with −0.02 (neutral), indicating that, compared to other modalities, it is not very useful for efficient task performance. Still, Tactile had the second highest Hedonic Quality with 1.47. Overall, Tactile was demanding and achieved rather mixed results.

Auditory had the best NASA TLX score across all four categories. Especially its Mental Demand score was approximately 2.4 times better (lower) than for Baseline and Tactile. Moreover, it had the highest Pragmatic Quality with 1.91 but, in contrast, the lowest Hedonic Quality with 0.84, resulting in an overall score of 1.38 that is slightly lower than Visual. Apart from this, Auditory performed very well across all of our measurements and like Visual, it yielded a positive result (i.e., > 0.8) in Pragmatic and Hedonic Quality.

4.4 Ranking
Auditory and Visual were each ranked 1st by half of our participants. Baseline and Tactile often received the last rank. Especially Baseline only once received rank two and was ranked last by nine out of 16 participants. Tactile was ranked 2nd by five participants, placing it between Auditory (six times) and Visual (four times) for that rank. Overall, Auditory and Visual outranked the other two conditions, while Tactile was also ranked better than Baseline.
5 DISCUSSION

In the following, we discuss our most interesting findings. Baseline expectedly resulted in more notifications missed (14% total) and less confidence compared to the other approaches. However, it is surprising that half of our participants were unclear whether they received all notifications in this condition. This uncertainty might lead to a loss of concentration while presenting. Additionally, almost every participant stated being overwhelmed and frustrated without additional help of the wrist-worn device. One participant noted a high mental demand (“It was confusing and demanding to be aware of everything”, P9). Overall, about 20% of participants believed that their response was slow without supplemental notifications.

The results obtained for the Auditory modality indicate the effectiveness of using channels other than the occupied visual channel. However, using auditory cues might not work in every setting. It is noteworthy that despite the similarities in terms of cognitive load, there is a difference between Auditory and Visual in terms of mental demand according to the NASA TLX scores.

Furthermore, Tactile proved to be mentally demanding, which may be explained by the complexity of the vibration patterns. However, our participants were still able to process the vibration stimuli received under a high workload (“I wasn’t afraid to miss any important information”, P2). Future research might investigate different haptic patterns. Especially, the lower Pragmatic Quality (-0.02) in combination with the relatively high Hedonic Quality (1.47) might indicate that other patterns might be worth investigating. Our participants stated that it was complicated to parse the incoming vibration pattern (“It was very difficult to keep in mind the meaning of each pattern”, P12), and thus they used the vibration rather as an indicator to look for a new notification on screen (“It is not necessary to look at the wristband. You look at the Zoom window quite quickly when the vibration is active”, P8). Searching the screen turned out to be more convenient for some users than trying to understand the haptic pattern. Participants, therefore, suggested to combine Tactile with Visual. With this approach, vibrations would grab the user’s attention, while the color would then convey the needed semantics, splitting up “signal” and “message” of a notification. Choosing Visual over Auditory for these suggestions might also have occurred due to Visual having achieved the highest Hedonic Quality among all three modalities. Our study has shown that even though the visual channel of a presenter in a mentally demanding task might be very occupied, peripheral illumination is still perceivable with accuracy.

Overall, we could observe that modalities with lower cognitive load received better placements in the final ranking. Especially in presentation settings, it is important to disturb the presenter as little as possible. Using Visual or Auditory modalities met these criteria.

6 LIMITATIONS

We set up our study to resemble a common situation for presenters giving online presentations. Still, our participants gave presentations without a real audience apart from the study conductor. Therefore, notification recognition accuracy and the respective confidence might change in a fully realistic presentation scenario. Additionally, the participants not being familiar with the presentations beforehand could have lowered the level of cognitive load on each participant as they might have simply read the presenter notes out loud. In our study, we also only used one modality at a time, to gain an understanding of how this unimodal feedback performs in a presentation scenario. However, combining multiple feedback modalities might work well for these scenarios. Lazaro et al. [11] suggest, for example, to use auditory signals to attract users’ attention and then provide visual signals for confirmation. We saw a similar trend with our participants suggesting to use vibrotactile feedback to attract their attention. One disadvantage of our approach is that it uses another device besides the laptop. Future studies could investigate whether smart watches already can be used for the feedback variants. Since we only added notifications as additional cues for the participants, it is hard to tell the efficiency of our prototype in regards to conveying semantic information to the participants alone. Finally, NASA TLX measures the total
mental demand, i.e., of primary and secondary task combined. As a result, it only allowed us to compare the combined mental demand, not how much each task contributed to it.

7 CONCLUSION AND FUTURE WORK

In this work, we investigated notifications via different modalities for presenters in online presentations using a wrist-worn prototype. For this, we used visual, auditory, and vibrotactile feedback. Our results indicate that visual and auditory cues can be used confidently if users need to deduce a specific meaning (e.g., by identifying a color and matching it to a notification type). Still, using auditory feedback for presentations highly depends on the individual presentation setting and will not be suitable in all situations. Depending on the scenario, only one of the modalities we tested might be feasible. For example, acoustic notifications might also disturb the talk for the audience. Vibrotactile feedback caused more confusion among participants and is rather suitable when it is already sufficient to notify the user that some notifications happened. Overall, a combination of multiple modalities might yield a very good user experience. Vibrotactile feedback could make the user aware of an important notification, while visual or auditory cues then make it easier to understand the exact message. Nowadays, smart watches might be a good place to include this feedback, and therefore help to improve (online) presentations, e.g., by displaying emojis.

Since our study setup did not include a live audience, adding this is a logical next step to verify our findings in the field. While we focused on online presentations, in-person presentations might also benefit from the kind of presenter notifications we studied. Many kinds of notifications would be equally relevant to online and face-to-face presentations. For example, presenters could receive information on the time remaining, or on the overall fatigue level of their audience, to adapt their presentation accordingly. This could increase the quality of both face-to-face and online presentations.

Overall, we hope that our work can foster further research in the field of online presentations, which are becoming increasingly important in our world, and that our findings will also be of use beyond the current pandemic.

ACKNOWLEDGMENTS

This project was funded in part by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, R. Schäfer under grant no. 42586911 to Jan Borchers, and T. Wagner under grant no. 425867974 to Anke Huckauf and Enrico Rulzio), as part of Priority Program SPP2199, Scalable Interaction Paradigms for Pervasive Computing Environments.

REFERENCES