Exploring Audio, Visual, and Tactile Cues for Synchronous Remote Assistance

Sebastian Günther
TU Darmstadt
Darmstadt, Germany
guenther@tk.tu-darmstadt.de

Sven Kratz
Harman International
Mountain View, CA, USA
sven.kratz@harman.com

Daniel Avrahami
FXPAL
Palo Alto, CA, USA
avrahami@fxpal.com

Max Mühlhäuser
TU Darmstadt
Darmstadt, Germany
max@informatik.tu-darmstadt.de

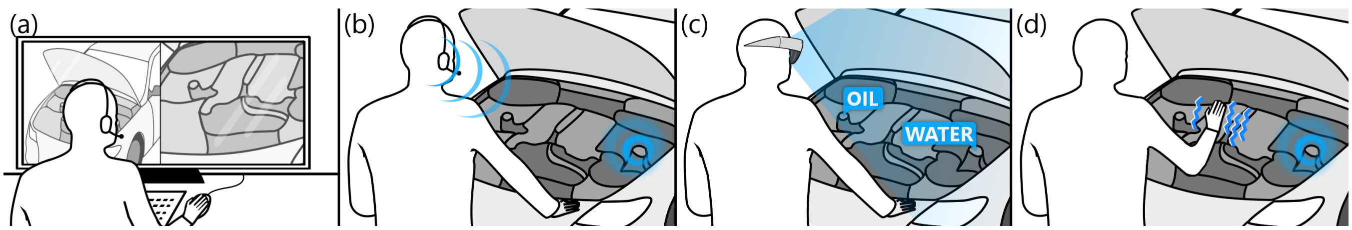

Figure 1: Example remote assistance scenario for using audio, visual, and tactile cues: (a) A remotely located expert supports a worker in the field through (b) audio, (c) visual, and (d) vibrotactile cues.

ABSTRACT

Today, remote collaboration techniques between field workers and remotely located experts mainly focus on traditional communication channels, such as voice- or video-conferencing. Those systems may not be suitable in every situation or the communication gets cumbersome if both parties do not share a common ground. In this paper, we explore three supporting communication channels based on audio, visual, and tactile cues. We built a prototypical application implementing those cues and evaluated them in a user study. Based on the user feedback, we report first insights for building remote assistance systems utilizing additional cues.

CCS CONCEPTS

• **Human-centered computing** → **User studies**; *HCI theory, concepts and models*;

KEYWORDS

Vibrotactile Feedback, Haptics, 3D-Space, Navigation, Spatial Guidance, Assistive Technology, Augmented Reality, Audio Cues, Remote Collaboration

PETRA '18, June 26–29, 2018, Corfu, Greece

ACM ISBN 978-1-4503-6390-7/18/06...$15.00
https://doi.org/10.1145/3197768.3201568

ACM Reference Format:
Sebastian Günther, Sven Kratz, Daniel Avrahami, and Max Mühlhäuser. 2018. Exploring Audio, Visual, and Tactile Cues for Synchronous Remote Assistance. In *PETRA '18: The 11th PErvasive Technologies Related to Assistive Environments Conference, June 26–29, 2018, Corfu, Greece.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3197768.3201568

1 INTRODUCTION

Nowadays, persons who need assistance in completing a complex task often rely on experts that are physically present. Typically, such assistance is practical if the person in need is not experienced or can only succeed with additional advice. An expert visiting the person on site can be a solution in situations, such as a car breakdown where the driver acts as the unqualified worker during an accident and a professional repair mechanic gives instructions as the expert. In a business environment, an exported machine could be maintained by a local field technician assisted by a remote certified expert. In a casual scenario, a traveler could get one-directional tourist information from friends at home. However, while highly effective to provide help colocated, it is also time-consuming to travel large distances, especially if the time to solve tasks is shorter than the time to travel. Further, traveling results in high costs for the involved parties. For instance, the US Travel Association reported that business travel in 2016 totaled in \$ 307.2 billion[1].

To tackle such high costs and travel times, remote assistance systems emerge to provide support over a distance. Traditional, those systems are based on voice-only or video communication, such as phone or video conferencing tools. Hence, they are limited within their scope of verbal describing or pointing at objects (e.g., "*look at*

[1] https://www.ustravel.org/answersheet, last accessed 29/03/2018

	Audio	Visual	Tactile
Head	stereo sound for view direction	markers on screen edges	vibrations on headsides
Hand	n/a*	highlighting spots	vibration on hand
Body	surround sound for spatial orientation	embedded movement animations	multi-vibration patterns on body joints

Table 1: Possible cues for different body parts.
Guiding a hand with audio is probably inefficient. Though, they may accompany other cues.

this", "*reach there*"). This can be cumbersome if the parties do not share a common ground or have language constraints [5, 23]. Moreover, voice communication can interfere with noise or overload the worker's perception requiring a high amount of conscious attention while solving the actual task. Some more advanced systems suggest video-mediated approaches, where the field worker has a separate display showing instructions of the expert [1, 7, 18, 21]. Some of them are completely stationary with a fixed setup, while others add another degree of freedom by giving the expert extra control (i.e., movable camera [11]).

To overcome those limitations, the research community focused on supportive methods via 1) audio, 2) visual or 3) tactile communication channels. Each provides subtle context information and may reduce the need for verbal agreements. This context information may concern the head orientation, the hand motion or even the whole body movement (e.g., *walking somewhere* or *turning around*). Figure 1 shows an example of a car breakdown scenario and how a remote expert could assist a person with different communication approaches.

In this paper, we contribute a first prototype implementing three different cues: Audio, Visual, and Tactile. We evaluate these cues individually in an exploratory study to identify their benefits and limitations. Finally, we report the results based on the user feedback as lessons learned.

2 TYPES OF CUES AND BACKGROUND

There exists a large body of work focusing on individual supporting communication channels based on audio, visual and tactile cues. Those cues can be used to enhance traditional remote assistance systems by actuating various body regions or subtle steer the attention of users towards a highlighted object. In table 1, we show an example design space of how such cues can affect certain regions of the users, such as the head, hands or the whole body. Further, we give an overview of those types of supporting cues for remote assistance systems in the following.

Audio Cues. Besides voice-to-voice communication, basic audio signals can be used to transmit spatial information. For example, a basic audio cue can be thought of a subtle beeping sound. In this case, it is possible to use both audio channels of a stereo headset to move the attention and field of view either to the left or right. Prior work proposes to use spatial audio to assist navigation systems [13, 16, 24]. Audio cues are most basic to drive the remote worker but have the advantage that they do not need any extra hardware besides a headset. However, they also suffer from noisy environments.

Visual Cues. Video conferencing systems give a good notion about the overall scene, but describing objects only by speech can be cumbersome. This is especially the case if multiple objects are too similar or cannot be easily identified. A head-mounted camera can be used to capture the first-person view of a worker to increase the effectiveness [6, 15]. Further, a head-mounted display (HMD) can augment visible information on target objects in form of icons or annotations [8, 9], as well as virtually showing the hands of the expert performing the task [14, 25]. Two different approaches emerged: using Virtual or Augmented Reality systems. VR environments are similar to telepresence systems, where an expert shows the worker how to solve the task on the non-transparent HMD [2, 3], while Augmented Reality (AR) solutions use look-through displays and augment information directly onto the real-world. Combining HMD with an attached camera and AR, not only the worker's field of view can be shared, also the remote expert can provide additional information directly on objects.

Tactile Cues. Tactile cues aim to assist by driving a user's attention through vibration actuation or similar haptic feedback. This can be useful in scenarios where a system is barely visible or the worker can not directly look at a specific object. An expert can steer the user's attention without having the worker to visually focus on an object. For example, in a different work, we presented a vibrotactile glove for guiding persons towards non-visible targets in 3D space [10]. Similar approaches supporting close-range navigation are not limited to the hand, and can also be used for the arm [27, 28] or head [17], as well as navigating the users' whole body movement with no need to refocus the attention, such as full body navigation systems [12, 26].

2.1 Combining different cues

The presented types of cues can also be combined into single interfaces. Research, therefore, provides different work that compares them. For instance, Funk et al. [4] and Kosch et al. [19] explored how such cues can be used as feedback for error prevention or how they can be used for assisting persons with cognitive impairments.

While practical and useful, there is a lack of research done on using those cues for remotely located users. Therefore, in this paper, we conduct a user study exploring how a remote assistance system can be enhanced by implementing such cues and how users perceive them to cooperatively solve a task over a distance.

3 SYSTEM AND PROTOTYPE

To evaluate the concepts, we built two applications: 1) a client for the field worker and 2) an operator interface for the remote expert. Both are implemented with Unity3D and using the META 1 SDK. The worker client outputs the information given by the expert. The operator client is used to see the worker's perspective and to select the cues. For our prototype we implemented following cues:

(1) Audio cues utilizing stereo signals to guide a worker's head.
(2) Visual cues enabled by augmented labels on target objects.
(3) Tactile cues guiding a worker's hand with vibration motors.

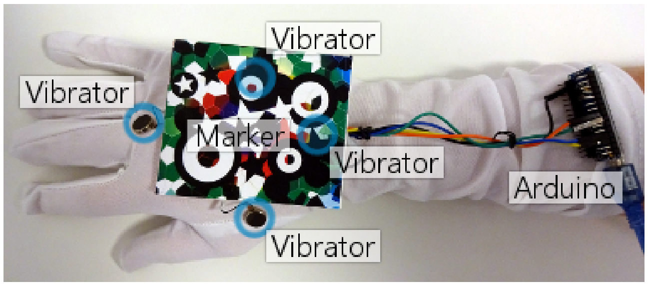

Figure 2: Our tactile glove prototype with four vibration actuators and the connected microcontroller. For tracking the glove's position, we added a visual marker.

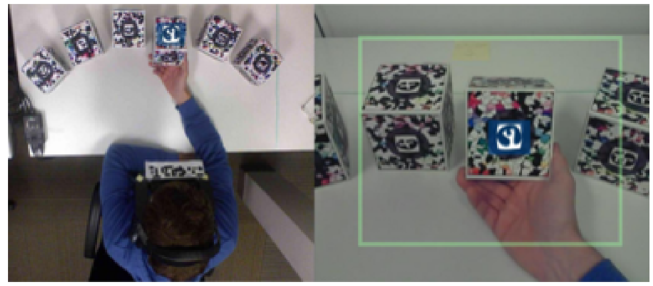

Figure 3: Remote expert application showing the overview perspective (left) and worker's HMD perspective (right) while completing a task with augmented labels.

3.1 Worker Client

The worker client displays the additional cues to the field worker and streams a live video feed to the remote expert. The visual cues are displayed with a head-mounted display (HMD). A headset is used for voice communication and audio cues. Tactile cues are realized by a tactile glove with four attached vibration motors.

To give the remote expert a supporting overview of the scene, we mounted a second camera above the workspace. This camera is also used to track a 2D position of the worker's hand, head and all objects. The HMD returns the head orientation with the built-in compass, whereas the built-in camera streams a live video feed to the remote expert.

Audio: Stereo Signals. The headset is used for speech, but also to output audio cues in form of stereo signals. Depending on the worker's head orientation, the left or right audio channel has a sound with a higher intensity. This means, if a highlighted object is left to a worker's view, the left audio channel is active, while the right one is muted and vice-versa. The closer the worker's view is towards the target object, the more intense the audio signals get. If the worker found the target object by directly looking at it, the signal changes its frequency and both audio channels are active.

Visual: Head-Mounted Display. To augment annotations directly within a worker's view, we used the Meta 1 glasses. This HMD device provides a transparent display and a camera integrated which is used to stream a live video feed of the worker's perspective to the remote expert. The HMD is tracked by a top-mounted camera to locate the worker's position while the built-in compass returns the precise head orientation.

However, the HMD has some technical limitations, since the internal display does not provide a large field of view (FOV) covering the whole FOV of the user.

Tactile: Glove. For tactile guidance, we mounted four small vibration motors onto a glove. An Arduino Nano was used to control pulsing vibration patterns [22]. The position is then mapped to a 2D coordinate system based on the top-mounted camera resembling a horizontal plane. Depending on the hand position and the target object, the vibration actuators change their intensity that the closest to the object vibrate, while others are disabled. Figure 2 shows the glove prototype and the location of the vibration actuators.

3.2 Operator Client

The remote expert works on a stationary desktop environment. The operator application has two main views to work with: 1) the live video feed of the worker's HMD and 2) an overview video feed of the remote workspace (see figure 3). The first-person video feed also visualizes the same information the worker is able to see through the HMD. The overview video shows the whole remote workspace and is planned to be replaced by an auto-generated 3D image. Both views can be used to select objects by clicking on them.

The current prototype uses previously scanned objects. Our tracking software returns the relative position of them and the identifiers for both, the overview and HMD video feed. If one or both views lost track of an object, the system tries to interpolate the actual position with the last known position.

4 USER STUDY

We conducted a qualitative study to compare the effectiveness and to gain first insights of the supporting cues. We recruited eight participants randomly grouped into four teams (Workers: W1-W4; Experts: E1-E4). Four of them were female and all were aged between 20 and 50. Each participant had none or minimal experience with HMDs or AR. Three of them use remote-support a few times a year, while the others are never used them or only a couple of times. However, two-thirds provide help to other persons using traditional remote assistance systems at least a few times per week.

4.1 Design and Task

We modified the available communication channels as the dependent variable, resulting in a total of four conditions: 1) voice-communication only, 2) audio signals, 3) augmented labels, and 4) tactile feedback. We counterbalanced the conditions using a Balanced Latin Square design and each participant had either the role of the remote expert or the field worker. In addition, the remote expert always had a live video feed showing the field worker's ego perspective and a top-down view.

We designed the task with simplicity in mind, together with low learn effects after each trial. It should be easy to understand what each side has to do, but in each trial, the participants should not be able to use knowledge from a previous trial. Therefore, we decided to design a sorting task of six similar looking boxes which had to be

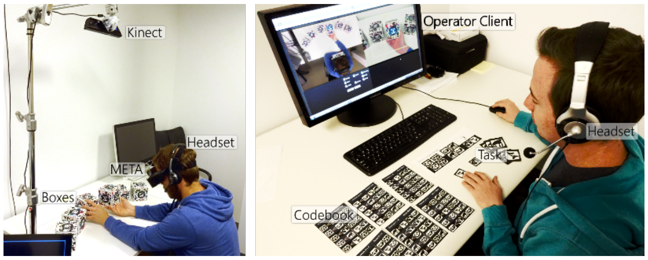

Figure 4: Setup of the study: a field worker wearing the HMD and a Kinect for a top-down view (left), and the remote expert using a workstation with the provided codebook (right).

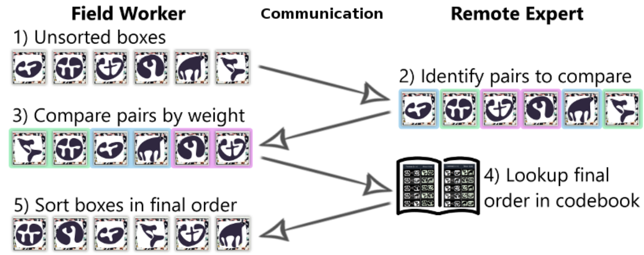

Figure 5: Sample procedure showing the responsibility of both, the field worker and remote expert.

placed in a correct order by the worker while the remote expert assists the identification process (cf. Figure 5.1). To get the correct sorting order, all boxes had to be compared by their weight (cf. Figure 5.2 and 5.3). Each had a different weight, starting at 100 grams and 200 grams heavier for each subsequent box. Informal pre-tests showed that a difference of 200 grams was easily distinguishable for users. In addition, the boxes had an abstract, not easily describable symbol printed on them. This was intended to be hard to communicate by voice only, but still possible to identify by the participants.

The task was to always compare three predefined pairs of boxes and decide which box is heavier. To actively enforce both participants to communicate, the remote expert had no knowledge of the actual weight of a box, while the worker never knew which boxes to compare until communicated by the expert. Once the given pairs were put into correct order by the field worker, they had to show the results to the expert.

In a next step, the remote expert had to look a final sorting order based on the previous results. The correct order had to be identified in a simple codebook with the help of the symbols printed on the boxes (cf. Figure 5.4). Once the final order based on the code was identified, the expert had to tell the remote worker how to reorganize the boxes in a last step (cf. Figure 5.5). If both sides agreed to have successfully solved the task, the trial was completed.

Even though the task was abstract, it resembles the real world communication in remote assistance scenarios, such as health emergencies. Before a medical expert (remote operator) is able to give a diagnosis, contextual feedback of the paramedic (field worker) is necessary (i.e., measure blood-pressure).

4.2 Setup

We located the expert in a room with a stationary desktop computer. There was a headset prepared for the participant as well as the codebook for finding the correct sorting order. At the beginning of each trial, we handed a sheet with the target boxes to the remote expert.

The remote worker was located in a second room. We connected the HMD glasses and a stereo headset to a desktop PC running our system. A second camera was mounted above the workplace to record the whole scene. For the task, the six boxes with random patterns and different weights are placed on the desk.

The remote expert and worker were spatially separated in different rooms and only connected over a local network. Both sides could only communicate with the given channels depending on the current condition. In each trial, we logged the time the participants needed to solve the puzzle. For further evaluation, we also recorded the overview screen of the operator application and saved all audio communication. Figure 4 shows the setup of the worker and remote expert workplaces.

4.3 Procedure

After welcoming the participants and explaining the system, we introduced them to their respective roles. Before the first trial started, participants filled a demographic questionnaire. The remote expert had to wear the headset and sit down in front of the desk, while the field worker also had to wear the HMD. If the trial was the tactile condition, the remote worker also had to wear the vibration glove.

Once the participants were ready, each trial started by giving the expert the current task which explained which boxes have to be compared (cf. Figures 5.1 and 5.2). Then, the participants were asked to work together over the available communication channels depending on the current condition (voice-only, audio, visual, tactile). After they identified the box pairs, the remote worker should tell the expert the compared weights of the boxes (cf. Figure 5.3). Then, the expert could get the final order of the boxes with the help of the provided codebook (cf. Figure 5.4). Once the expert told the worker that order, they had to tell that they completed the task and we logged the Task Completion Time (TCT).

Both, the expert and worker, then had to fill out questionnaires with regards to their role and the current task. In this intermediate questionnaires, we used a Likert scale for the subjective ability to solve the task, the communication simplicity, and the effectiveness and distraction of the supporting cue. Moreover, we wanted to know how each side rated their counterpart. This included the helpfulness of the other person, and how effective or disruptive they had been. The participants could then take a short break and once they were ready, the participants continued with the next condition. In Figure 5, we show the whole procedure for a sample trial and the responsibilities for each side.

After completing the task for all four conditions, the participants filled out a final post-questionnaire where they had to rank each condition by their perceived effectiveness, disturbance and their overall rating. In total, the experiment took about 60 minutes per session.

4.4 Results

In the following, we present the results of our pilot study. Due to the low number of participants, we did not consider tests for statistical significance and focus on the qualitative feedback. Further, we report the means and standard deviations.

All participants agreed that the supporting channels have potential to enrich the communication and to assist them in remote collaborative tasks. Regarding their preferred communication channels, the questionnaires showed visual as the most and audio signals as the least preferred and effective way to communicate. However, with regards to the TCT, audio was the fastest (voice: M = 2:55 min, audio: M = 2:29 min). Labels and tactile cues were contemporaneous equal (visual: M = 3:13 min, tactile: M = 3:14 min).

The field workers answered that they could solve the task most easily with labels (voice: M = 4.25, SD = 0.8; audio: M = 3.5, SD = 1.5; labels: M = 4.75, SD = 0.4; tactile: M = 4.25, SD = 0.8). Even though visual labels were found most assisting, the distractiveness of the provided communication channels were similar for each condition (audio: M = 3.75, SD = 0.8; visual: M = 3.25, SD = 0.8; tactile: M = 3.75, SD = 0.8). Participants replied that visual labels tend to disturb the workers more in completing the task if the remote expert does not disable the overlay after objects are identified.

Both sides showed different feelings regarding their effectiveness as a team. While the remote experts thought they have been an effective team for every condition (voice: M = 4.75, SD = 0.4; audio: M = 4.75, SD = 0.4; visual: M = 4.5, SD = 0.5; tactile: M = 4.0, SD = 1.2), the workers were insecure for voice-only (voice: M = 3.75, SD = 1.6; audio: M = 4, SD = 1.2; visual: M = 4.5, SD = 0.5; tactile: M = 4.25, SD = 0.8). Overall, the participants felt the most effective as a team while solving the task with augmented labels, whereas audio was the least (voice: M = 3, SD = ; audio: M = 1.5, labels: M = 3.5, tactile: M = 2.75).

The design aimed to be easy understandable and not possible to solve with knowledge of earlier trials. Thus, learning effects should have no impact. However, we observed that the TCT of a trial was faster for subsequent trials in almost every case. Teams subconsciously started to identify boxes by naming the abstract symbols and rapidly had a common ground [5]. For example, one symbol reminded the participants of a *fish* or *dolphin*, while other symbols were described as a *mouth* or *house*.

4.4.1 Audio Cues.

Audio cues were the least liked supporting communication cue by the workers (worker: M = 2.5, experts: M = 4). However, the field workers appreciated the concept and told "*the sound was interesting and somewhat intuitive*" (W1). We observed that participants had problems to identify the correct object if they are too close to others. Interestingly, W3 reported it "*helped me to get to the right place more quickly than the verbal instructions*". The remote experts, however, described audio cues as effective, e.g. E4 said "*audio seemed to help the worker a little bit more*".

4.4.2 Visual Cues.

Visual cues in form of augmented labels were the most liked channel (worker: M = 3.75, experts: M = 3.75). Both sides had the impression to be more effective. W3 liked "*being able to see the image needed to match without the urge to describe it*".

However, a big issue was that the HMD did not work well for people with a debility of sight. W2 was "*not able to use the AR because my eyes don't work*". Further, when overlaying the users' field of view with too many information, they may have issues focusing on the relevant. Hence, it is important to have a trade-off between the real world scene and number of total augmentations. W3 suggested that "*a brief flash*" of a label could have been sufficient enough rather than continuously highlighting it.

The remote experts thought similar as the field workers. E3 positively mentioned, that selecting "*multiple boxes helped*" for telling the worker which boxes to compare. E1 told that selecting the target boxes helped to be more effective because "*it took less time to describe which boxes to chose than talking*".

4.4.3 Tactile Cues.

The tactile cues had a mixed reception (workers: M = 3, experts: M = 3.75). In general, the participants found the concept interesting and highlighted its novelty. W3 appreciated "*the concept in general [..] for guidance*". W2 explained that while the vibration is useful, it should automatically suppress once the target object was found because the expert sometimes forgot to turn it off. However, W4 told that more time to learn the vibration patterns was needed and the precision could be increased. Further, we observed that all workers hovered over or touched a box and asked the expert if they are correct.

The remote experts agreed that tactile accuracy for guiding the hand of a worker is promising. However, E1 saw more potential with regards to the accuracy ("*The accuracy is not good, so I gave up sending tactile feedback to a worker from the middle of this task*"). E4 also had the impression that signals "*seemed to be a little bit delayed*". However, the actual delay was always below a second.

4.4.4 Verbal Communication.

Comparing verbal communication, we observed that augmented labels needed fewer words to describe and identify tracked objects. The remote experts could highlight objects faster without describing the worker explicitly what they mean. Contrary, audio signals and tactile feedback did not show this effect because the field workers often described the operator what they are hearing ("*i hear a sound on my left now*") or feeling ("*it is guiding me to this box, is this correct?*"). Similar to Kraut et al. [20], we observed that additional cues changed the participants' used dictionary. If no supporting cue was active, the participants had to describe and acknowledge what they mean (e.g., E: "*use the box with the weird fish symbol on it*", W: "*This one?*", E: "*No, the left.*", W: "*Okay, this!*", E: "*Correct.*"). With visual indicators, experts often only intervened if the worker reached for the wrong object.

5 LESSONS LEARNED AND DISCUSSION

Based on the results of our exploratory user study, we learned the following lessons for designing a remote assistance system with supporting cues.

Always provide voice communication as a base channel. Users are very effective in finding a common ground (e.g., [5]). Even though supporting cues can reduce the necessary amount of spoken words, speech is still powerful and helps users to locate objects precisely.

Provide visual cues to ease the communication. When augmenting visual labels and instructions directly in a worker's field of view, both sides can identify objects fast and may able to reduce cognitive load by reducing the amount of verbal communication.

However, visual cues should not overlay the content. Some workers explained that visual cues can be disturbing if they do not disappear after they identified a correct object. This was mostly the case when the remote expert forgot to hide it.

Use tactile and audio cues as auxiliary cues. Both are suitable to assist the communication and can especially be useful if visual cues are not available due to environmental limitations. Further, audio and tactile feedback share similar advantages and can support situations where workers have to identify objects that are barely visible or cannot be visually focused. Similar to visual cues, they should suppress once a target is located. While both had benefits to solve the tasks, they could get distracting for workers if the expert forgot to disable them. As consequence, participants reported that it probably would be less disturbing if the cues would automatically suppress once a target was identified.

6 CONCLUSION AND FUTURE WORK

In this paper, we presented a prototypical remote assistance system for field workers. We contributed three possible interaction concepts for audio, visual, and tactile cues. In an exploratory user study, we evaluated those cues and contribute our lessons learned. Our results show that participants appreciated the auxiliary communication cues (audio, visual, and tactile) and their great potential for future collaborative tools for remotely located workers.

While visual cues in form of augmented labels improved their communication by identifying objects faster while reducing the number of spoken words, participants thought that tactile and audio cues are interesting concepts improving communication but need to be more precise. Hence, we want to elaborate how isolated cues can be further improved, especially in more realistic scenarios, such as a car-repair with noisy environments. Therefore, we also want to find out how other constraints affect the team work through additional communication regulations and rules. Further, we want to explore how combinations of multiple cues assist during collaborative tasks and how a system can select cues depending on the context of a task.

7 ACKNOWLEDGEMENTS

We thank the FX Palo Alto Laboratory, Inc. (FXPAL) in Palo Alto, California, USA for their valuable support. This work was supported by the German Federal Ministry of Education and Research (BMBF) SWC "ARkTiP" (01IS17050).

REFERENCES

[1] L. Alem and J. Li. 2011. A Study of Gestures in a Video-mediated Collaborative Assembly Task. *Adv. in Hum.-Comp. Int.* 2011, Article 1 (Jan. 2011), 7 pages. https://doi.org/10.1155/2011/987830

[2] J. Amores, X. Benavides, and P. Maes. 2015. ShowMe: A Remote Collaboration System That Supports Immersive Gestural Communication. In *Proc. CHI EA '15*. ACM, New York, NY, USA, 1343–1348. https://doi.org/10.1145/2702613.2732927

[3] X. Benavides, J. Amores, and P. Maes. 2015. Remot-IO: A System for Reaching into the Environment of a Remote Collaborator. In *Adj. Proc. UIST '15*. ACM, New York, NY, USA, 99–100. https://doi.org/10.1145/2815585.2815738

[4] M. Funk, J. Heusler, E. Akcay, K. Weiland, and A. Schmidt. 2016. Haptic, Auditory, or Visual?: Towards Optimal Error Feedback at Manual Assembly Workplaces. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments - PETRA '16*. ACM Press, New York, New York, USA, 1–6. https://doi.org/10.1145/2910674.2910683

[5] S. Fussell and R. Krauss. 1989. The effects of intended audience on message production and comprehension: Reference in a common ground framework. *Journal of Experimental Social Psychology* 25, 3 (1989), 203–219. https://doi.org/10.1016/0022-1031(89)90019-X

[6] S. Fussell, L. Setlock, and R. Kraut. 2003. Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. In *Proc. CHI '03*. ACM Press, New York, New York, USA, 513. https://doi.org/10.1145/642611.642701

[7] S. Fussell, L. Setlock, J. Yang, J. Ou, E. Mauer, and A. Kramer. 2004. Gestures over Video Streams to Support Remote Collaboration on Physical Tasks. *Hum.-Comput. Interact.* 19, 3 (Sept. 2004), 273–309. https://doi.org/10.1207/s15327051hci1903_3

[8] S. Gauglitz, B. Nuernberger, M. Turk, and T. Höllerer. 2014. In Touch with the Remote World: Remote Collaboration with Augmented Reality Drawings and Virtual Navigation. In *Proc. of VRST '14*. ACM, New York, NY, USA, 197–205. https://doi.org/10.1145/2671015.2671016

[9] S. Gauglitz, B. Nuernberger, M. Turk, and T. Höllerer. 2014. World-stabilized annotations and virtual scene navigation for remote collaboration. In *Proc. UIST '14*. ACM Press, New York, New York, USA, 449–459. https://doi.org/10.1145/2642918.2647372

[10] S. Günther, F. Müller, M. Funk, J. Kirchner, N. Dezfuli, and M. Mühlhäuser. 2018. TactileGlove: Assistive Spatial Guidance in 3D Space through Vibrotactile Navigation. In *Proceedings of the 11th ACM International Conference on PErvasive Technologies Related to Assistive Environments - PETRA'18*. ACM.

[11] P. Gurevich, J. Lanir, B. Cohen, and R. Stone. 2012. TeleAdvisor: A Versatile Augmented Reality Tool for Remote Assistance. In *Proc. CHI '12*. ACM, New York, NY, USA, 619–622. https://doi.org/10.1145/2207676.2207763

[12] W. Heuten, N. Henze, S. Boll, and M. Pielot. 2008. Tactile Wayfinder. In *Proc. NordiCHI '08*. ACM Press, New York, New York, USA, 172. https://doi.org/10.1145/1463160.1463179

[13] S. Holland, R. D. Morse, and H. Gedenryd. 2002. AudioGPS: Spatial Audio Navigation with a Minimal Attention Interface. *Personal and Ubiquitous Computing* 6, 4 (2002), 253–259. https://doi.org/10.1007/s007790200025

[14] W. Huang and L. Alem. 2011. Supporting Hand Gestures in Mobile Remote Collaboration: A Usability Evaluation. In *Proc. BCS-HCI '11*. BCS, Swinton, UK, 211–216. http://dl.acm.org/citation.cfm?id=2305316.2305356

[15] S. Kasahara and J. Rekimoto. 2015. JackIn Head: Immersive Visual Telepresence System with Omnidirectional Wearable Camera for Remote Collaboration. In *VRST '15*. ACM, NY, USA, 217–225. https://doi.org/10.1145/2821592.2821608

[16] B. Katz, S. Kammoun, G. Parseihian, O. Gutierrez, A. Brilhault, M. Auvray, P. Truillet, M. Denis, S. Thorpe, and C. Jouffrais. 2012. NAVIG: augmented reality guidance system for the visually impaired. *Virtual Reality* 16, 4 (2012), 253–269. https://doi.org/10.1007/s10055-012-0213-6

[17] O. B. Kaul and M. Rohs. 2017. HapticHead: A Spherical Vibrotactile Grid Around the Head for 3D Guidance in Virtual and Augmented Reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 3729–3740. https://doi.org/10.1145/3025453.3025684

[18] D. Kirk and D. Stanton Fraser. 2006. Comparing remote gesture technologies for supporting collaborative physical tasks. In *CHI '06*. ACM Press, New York, NY, USA, 1191. https://doi.org/10.1145/1124772.1124951

[19] T. Kosch, R. Kettner, M. Funk, and A. Schmidt. 2016. Comparing Tactile, Auditory, and Visual Assembly Error-Feedback for Workers with Cognitive Impairments. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, 53–60.

[20] R. Kraut, S. Fussell, and J. Siegel. 2003. Visual Information As a Conversational Resource in Collaborative Physical Tasks. *Hum.-Comput. Interact.* 18, 1 (2003), 13–49. https://doi.org/10.1207/S15327051HCI1812{_}2

[21] H. Kuzuoka. 1992. Spatial workspace collaboration. In *Proc. of CHI '92*. ACM Press, New York, New York, USA, 533–540. https://doi.org/10.1145/142750.142980

[22] S. Lee and T. Starner. 2010. BuzzWear. In *Proc. of CHI '10*. ACM Press, New York, New York, USA, 433. https://doi.org/10.1145/1753326.1753392

[23] R. Stalnaker. 2002. Common Ground. *Linguistics and Philosophy* 25, 5–6 (2002), 701–721. https://doi.org/10.1023/A:1020867916902

[24] V. Sundareswaran, K. Wang, S. Chen, R. Behringer, J. McGee, C. Tam, and P. Zahorik. 2003. 3D Audio Augmented Reality: Implementation and Experiments. In *Proc. of ISMAR '03*. IEEE Computer Society, Washington, DC, USA, 296–. http://dl.acm.org/citation.cfm?id=946248.946841

[25] F. Tecchia, L. Alem, and W. Huang. 2012. 3D helping hands. In *VRCAI '12*. ACM Press, New York, NY, USA, 323. https://doi.org/10.1145/2407516.2407590

[26] K. Tsukada and M. Yasumura. 2004. *ActiveBelt: Belt-Type Wearable Tactile Display for Directional Navigation*. Springer Berlin Heidelberg, Berlin, Heidelberg, 384–399. https://doi.org/10.1007/978-3-540-30119-6_23

[27] J. Zelek. 2005. Seeing by touch (haptics) for wayfinding. *International Congress Series* 1282 (2005), 1108 – 1112. https://doi.org/10.1016/j.ics.2005.06.002

[28] J. Zelek, S. Bromley, D. Asmar, and D. Thompson. 2003. A haptic glove as a tactile-vision sensory substitution for wayfinding. *Journal of Visual Impairment and Blindness* 97, 10 (2003), 621–632. http://www.afb.org/jvib/