MAPVI: Meeting Accessibility for Persons with Visual Impairments

Sebastian Günther[1], **Reinhard Koutny**[2], **Naina Dhingra**[3], **Markus Funk**[1], **Christian Hirt**[3], **Klaus Miesenberger**[2], **Max Mühlhäuser**[1], **Andreas Kunz**[3]
[1]TU Darmstadt, [2]JKU Linz, [3]ETH Zurich
[1]lastname@tk.tu-darmstadt.de, [2]firstname.lastname@jku.at, [3]kunz, hirtc, ndhingra@iwf.mavt.ethz.ch

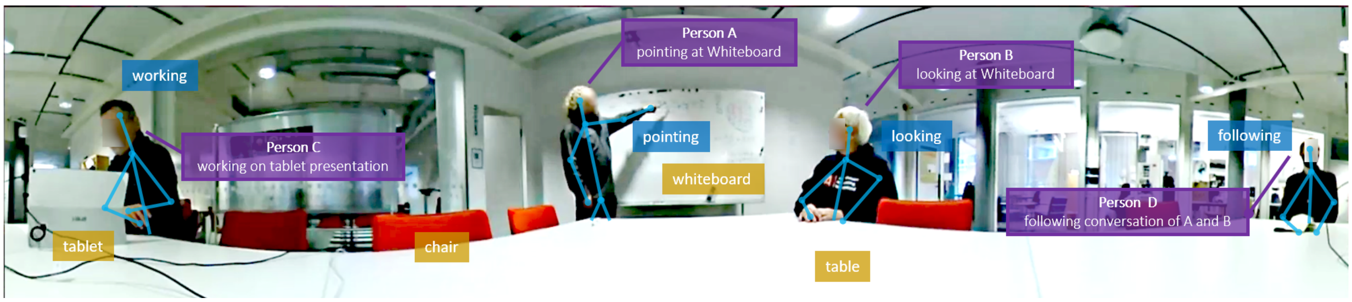

Figure 1: Vision picture of how the environment is captured and interpreted. Purple labels indicate other participants, while the blue labels show their sensed actions. Further, the skeleton tracking indicates the posture of each participant which is tracked through multiple cameras and sensors to improve the context quality. Yellow labels indicate the static properties of the environment.

ABSTRACT

In recent years, the inclusion of persons with visual impairments (PVI) is taking tremendous steps, especially with regards to group meetings. However, a significant part of communication is conveyed through non-verbal communication which is commonly inaccessible, such as deictic pointing gestures or the mimics and body language of participants. In this vision paper, we present an overview of our project MAPVI. MAPVI proposes new technologies on making meetings more accessible for PVIs. Therefore, we explore which relevant information has to be tracked and how those can be sensed for the users. Finally, those captured information get translated into a multitude of haptic feedback to make them accessible.

PETRA '19, June 5–7, 2019, Rhodes, Greece

ACM ISBN 978-1-4503-6232-0/19/06…$15.00
https://doi.org/10.1145/3316782.3322747

CCS CONCEPTS

• **Human-centered computing → Accessibility technologies**; *Pointing*; *Gestural input*; • **Computing methodologies → Machine learning**.

KEYWORDS

Assistive Technologies, Meetings, Haptics, Machine Learning

ACM Reference Format:
Sebastian Günther[1], Reinhard Koutny[2], Naina Dhingra[3], Markus Funk[1], Christian Hirt[3], Klaus Miesenberger[2], Max Mühlhäuser[1], Andreas Kunz[3]. 2019. MAPVI: Meeting Accessibility for Persons with Visual Impairments. In *The 12th PErvasive Technologies Related to Assistive Environments Conference (PETRA '19), June 5–7, 2019, Rhodes, Greece.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3316782.3322747

1 INTRODUCTION

Teamwork is characterized by intense collaboration which commonly work on a topic and solve given tasks. To support such teamwork, meeting rooms are usually equipped with analog whiteboards, flipcharts, sketching tables, etc. to enable, structure and document lively discussions. Within such discussions, information is made explicit using these tools while other information remains implicit, such as body

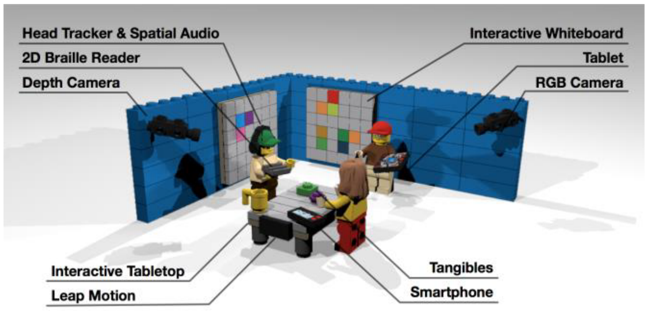

Figure 2: Conceptual setup of our meeting environment.

language, or the position of artifacts within the room. Consequently, such meetings cause inaccessibility issues for persons with visual impairments (PVI), in particular, for this implicit information.

Furthermore, such a lack of information for PVIs is not limited to in-situ meetings. During remote meetings, e.g., via video conferencing, it can also be challenging for PVIs to determine what the other participants are doing, for example if they are still attentive or show things in the camera so that there is also a need to translate events accordingly. Also, similar situations, such as spontaneous group discussion at workplaces or industrial workshops, need to be considered.

While many tools do exist to make information explicit for PVIs, most of the non-verbal information is still inaccessible to them. Thus, they lack important information and experience barriers which prevent their equal participation in team meetings which heavily rely on non-verbal communication (NVC) elements for coordinating and contributing tasks, and also for explanatory purposes [2, 31].

In mixed teams of visually impaired and sighted users, sighted users ought to make all the implicit information explicit by verbally expressing it. However, this imposes an additional workload on the sighted users and slows down the overall work process. Moreover, this requires a stern discipline of the sighted members, but the process of consciously expressing implicit information is frequently forgotten during a meeting.

While this mismatch of unconsciously not expressing NVC of sighted users and the available input channels of PVIs is a crucial problem, this is even exacerbated when looking at the artifact level. Here, information is represented in parallel on multiple horizontal and vertical information spaces (see Figure 2), resulting in a three-dimensional information space. In contrary, PVIs can only access information sequentially. Hence, we will address this by adding a layer of haptic actuations that translates the identified cues into tactile sensations, e.g., through vibration or thermal haptic feedback.

In this vision paper, we present our project *MAPVI*. In this project, we tackle the emerging challenges by enriching the perception of PVIs in meeting scenarios through capturing the meeting environments, reasoning the actions, and translating them into multi-modal haptic actuations.

2 RELATED WORK

For the scope of our MAPVI project, in the following, we sum up related approaches that focussed on interaction with meeting artifacts and projects focussing on non-verbal communication.

Interaction with Artifacts

Methods like Metaplan were originally designed for working with analog artifacts (e.g., cards), while IT-support was mostly limited to group decision support systems which only support binary tasks such as voting. First approaches in computer-supported cooperative work (CSCW) focused on shared-editing of documents. As stated by Dennis et al. [17], binary-tasks and shared-editings support were later merged into electronic meeting systems, and subsequently into systems for web-based collaboration. However, team working methods were not explicitly supported; rather, e.g., plain electronic blackboard functionalities were used.

While all the systems at the time using mouse and keyboard, LiveBoard [18] is one of the first systems that allows editing the artifacts directly using an optically tracked pen. In order to support the Metaplan method, Magerkurth & Prante [41] use PDAs to generate cards, which could be transferred to a shared electronic whiteboard, where they were rearranged and clustered. The PDAs communicate with the interactive vertical screen through the BEACH software [63] which is also able to address the i-Land environment [62], consisting of interactive walls, tables, and PDAs. This was further developed within the Stanford Interactive Workspaces Project [26]. A more recent approach to support the Metaplan method is introduced by Jetter et al. [25]. They use a back-projected interactive table together with a high-resolution vertical interactive screen. An overview of further augmented work environments is given by [12, 38].

Prior work was also done in making artifacts accessible to PVIs. An overview of older work is given in the MICOLE project [45]. More recently, Bourbakis [7] use a two-dimensional vibrotactile display which was used to show a three-dimensional environment to the user as an aid for navigation, while Brock & Kristensson [11] use a Kinect depth-sensing camera to sense a three-dimensional environment and to sonify approaching objects. However, sensing abstract information on interactive surfaces is still very limited. Another approach by Brock et al. [9] use a raised-line overlay on a touchscreen to output information to a PVI. Brewster & Brown 2004 [8] introduced tactile icons that helped PVIs to access structured content. While Braille displays are an established tool to access digital written content, they are obtrusive in particular during brainstorming meetings. Here, Shilkrot et al. [58] presented a finger reading device, which - worn like a ring - is able to sonify written text. A tactile map was developed by Brock [10], and an interactive workspace for helping PVIs to learn linear algebra was introduced by Almasri et al. [1]. More recently, TalkingCards were introduced [52], where information is conveyed via a tangible texture.

Beside the existence of IT-support for accessing digital artifacts, it is important to consider the mental model that PVIs could generate for the spatial layout of the information. The usual approach is to establish a second, virtual arrangement of artifacts that is (a) either a sequentialized representation to be rendered on established output devices (e.g. Braille displays, audio) or (b) to resemble the actual arrangement to a certain degree via more innovative modalities. This could, for example, be on a two-dimensional HyperBraille pad [48] which is augmented by audio notification or by adapting tangible interaction concepts as introduced by Kannengieser & Oppl [27]. However, this still does not result in a synchronized model for the spatial arrangement of artifacts and prevents the resolution of some NVC elements e.g. deictic gestures or spoken spatial references.

A more advanced approach is to retain the spatial relations and to facilitate the accessibility of the arrangement for PVIs as much as possible. This approach is desirable as it establishes a synchronized mental model for the spatial arrangement of artifacts among all participants, but it entails several issues and research questions [47] which are addressed in this paper.

One promising approach to provide accessible interfaces with an inherently synchronized mental model for the spatial arrangement is to use tangible objects on a digital and touch/object sensitive display, e.g. by Kannengieser & Oppl [27], with the PERCS system [14], or with Capricate [53]. Hence, the first systems [1] are on the way to enter the market. Here, the tangible objects are containers, similar to editable bubbles in mind-map tools, which can be associated with any information and manipulated on the display (e.g. arranged, connected, or grouped). The actions are tracked and integrated into the digital representation of the developed scenario.

To assess the level of synchronicity for the mental model of the spatial arrangement of artifacts, we have to determine the mental models that are established by PVIs when interacting. An initial approach was described by Kurniawan et al. [36], with the identification of a functional and a structural mental model established by the PVIs. A conceptual model for a software/hardware architecture supporting accessibility, in general, was introduced by Karpov & Ronzhin [28] and an assessment of the mental maps of spaces established by PVIs via haptic feedback is evaluated for synchronicity in [37].

[1] e.g. https://www.metasonic.de/touch

Non-verbal Communication

NVC is crucial for teamwork efficiency and can intuitively be understood by sighted users. Perception of body postures of others and the ability to interpret them as a body language leads to insights on the person's emotions and attitude, allowing more effective communication [32]. As originally stated by Mehrabian [44] and reaffirmed in subsequent publications [31], postures can be described along with an open-closed dimension which provides or denies access to body spaces. The former is done by opening up to others and the latter by hiding behind their self-imposed barriers. This implied attitude has a consequence for our willingness to engage in co-present collaboration as a group [42].

As identified by McDaniel et al. [43] in mixed focus groups and prioritized with an online survey among PVIs, the number, identity and position of participants as well as their facial expressions, body postures and hand gestures are the most helpful NVC elements for PVIs. Calvo & D'Mello [13] studied the combination of physiology, face, voice, text, body language, and complex multi-modal characterizations, whereas Zeng et al. [64] focused on modalities combined with facial expression recognition based on computer vision technologies, i.e. audio-visual fusion [51], [55], linguistic and paralinguistic fusion of facial expressions, head movement, and body gestures.

However, in order to convey this information to PVIs, it is the computer that must understand these NVC elements first. In this context, a clear distinction between 'on-surface', 'above-surface' and 'in-room' gestures is helpful. An approach to address some of the points above was made by Tan et al. [57]. Their project investigated a system that hints PVIs on approachability of strangers if they need assistance. For above-surface gesture-based user interfaces, Hilliges et al. [23] presented a vision-based tabletop that is capable of seeing hands and fingers and interpreting them as in-air gestures. This could also be adapted for tracking NVCs and making their meaning explicit. Similarly, Banerjee et al. [5] presented an in-air interaction technique to manipulate out-of-reach objects on tabletops. Müller et al. [46] investigated direct and indirect foot-based interaction with virtual contents. Klompmaker et al. [30] demonstrated an interaction technique which uses mobile phones for performing above-surface in-air interactions to manipulate spatially distributed artifacts, while Rader et al. [50] uses mobile phones for performing personalized above-surface interactions in collaborative tasks. De Araújo et al. [15] presented Mockup Builder, a combined on-surface / above-surface technique which uses motion sensors, a 3D projector, and the shutter glasses. It supports direct in-air interaction for sketching three-dimensional models on- and above-surface. The issue of capturing NVCs for conveying them to PVIs is, of course, the subject of the IT-based inclusion research community. McDaniel et al. [43] uses computer vision to detect faces looking at the PVIs and outputs this information on a haptic belt with vibrating actuators. This approach of using information from the physical world and translating it into haptic stimuli, was used by many projects to encode information into a haptic channel - not solely for PVIs: haptic information was used to guide PVIs with a leashed drone [4], guide persons towards invisible 3D targets [20], support remote collaboration [19], or tell travelers directions of points of interest [22].

Another system by Krishna et al. [33] uses computer vision to detect faces and facial expressions and sonifies the results to the PVIs. However, the systems discussed here require special equipment attached to the sighted or visually impaired user, which makes them less suitable for brainstorming meetings due to the preparations required. Moreover, none of the systems preserve the relation between NVCs and artifacts.

While sighted persons are clearly able to process these in real-time and focus on the most relevant aspects, a one-by-one translation of this information for the PVIs would lead to an overload of the cognitive capacity as they are mainly restricted to one-dimensional modalities.

Finally, another aspect of using assistive technology is its social implications. While assistive technology obviously first needs to fulfill its purpose of providing support for visually impaired persons, it is important that social implications are considered when using the assistive technology [3, 49].

One approach to improve the overall detection reliability and avoid false notifications for PVIs is the application of iterative fusion operator trees [60]. In this context, it is worth mentioning that the data fusion literature emphasizes the importance of developing novel data fusion algorithms in order to fully exploit this processing stage in applications where data fusion is involved [6, 56].

3 ACCESSIBLE MEETINGS CONCEPTS

In order to overcome the presented challenges, we define three areas which address the problem-space and are conceptualized in the following: 1) Capturing the environment, 2) Sensing and Reasoning the information, and 3) Translating them into haptic actuations.

Capturing

Using multiple surfaces to externalize information requires the functionality of moving or copying information from one workspace to another [62]. This generates a volatile, three-dimensional data space, which has to be captured and translated to PVIs.

In a loosely moderated Metaplan session, users will move between different interactive surfaces, addressing, browsing and modifying spatially distributed information [54]. Accessing this information (*who has changed which information where?*) is important for the participation of PVIs. Further, horizontal and vertical workspaces do not only have a different orientation but also have an impact on user behavior [35]. While tables are preferably used for information generation, vertical workspaces are used for information distribution and visualization. This may influence other variables like distance to the interactive surface, the accuracy of pointing gestures, choice of NVC elements, or chosen work methods. Consequently, a large variety of NVC elements have to be recognized, interpreted, and translated to PVIs.

Sensing and Reasoning

Modification of artifacts, augmented by NVC elements, constitute a meaning [29] that have to be derived, rated for relevance and delivered to PVIs with respect to their cognitive load. Further, multiple interactive surfaces also influence the data structure. For example, while on one single display, all data would be shown simultaneously, however, multiple displays lead to a thematic clustering of information, while also containing explicit or implicit cross-references. This makes it much more complicated to translate information to PVIs, since it may not be trivial to transform the information structure from multi-dimensional to a serial representation (e.g., due to loops caused by the cross-references in and between clusters).

While users on horizontal surfaces typically interact with the content using a pen or the fingers, multiple workspaces in different orientations will also trigger other gestures for interacting with the system, such as a grasping gesture to fetch useful information, and a releasing gesture to place the content on another workspace [62]. Thus, these gestures need to be unequivocally detected and correctly interpreted to guarantee a smooth interaction with the system.

In addition, data generated from various sources during meetings need to be stored, condensed and made accessible through queries. This information and the relations within need to be structured, hierarchized and modeled using ontologies for further processing and retrieval. Due to the sheer amount of information which is generated simultaneously from numerous sources, it needs to be condensed and, depending on multiple factors, tailored to the users' needs. Potentially important factors seem to be the users and their capabilities, the current situation within the meeting and preferences of the users, predefined ones as well as ones changed on the fly.

Actuating

Multiple interactive surfaces shape a spatially distributed information space and contain a larger amount of data than a single workspace. Thus, serializing all spatially distributed information on a one-dimensional user interface (Braille display) is not feasible any longer. As outlined in [47], new ways have to be researched on how a PVI can access, navigate and handle the increased amount of (spatially distributed) information. Nowadays, PVIs mainly use keyboards, cursor-routing, touch-gestures (e.g., [61]), or tangibles (e.g., [21]) to enter and manipulate information. While all are practical and useful in a lot of situations, the amount of distributed information increases and traditional interaction methods reach their limits. Therefore, to manage this big amount of distributed information, we will research new methods of interaction, e.g., to track and use in-air gestures of PVIs for orienting and positioning the input in the information space and for manipulating it (e.g., add, change, delete, etc.).

To encode the previously captured and interpreted information of such scenarios, we will investigate multiple types of non-obtrusive actuations that assist PVIs by enriching their awareness of their surroundings. Hereby, this can be reached through traditional auditive cues [59], or through subtle vibrotactile actuations guiding users attention towards a certain event or property (e.g., [19, 20]), or even through more active actuations, such as pneumatic feedback [16], or thermal haptic feedback [24].

4 PROJECT MAPVI

Our approach will start with an assessment of Metaplan sessions as an example of structured team meetings and group collaborations within an environment that comprises multiple surfaces, both, horizontal and the vertical. Based on this assessment, we expect answers on "how to capture" artifacts and NVCs and "how to make sense" to support the interpretation of inaccessible modalities for a presentation via accessible ones. Depending on the specific approach via a) established modalities or b) innovative modalities, the spatially distributed data from the NVCs and artifacts needs to be either de-spatialized or merely reorganized and translated in order to answer the question "how to actuate" the interpreted information. This will allow evaluating the combination of modalities which help in the establishment of synchronized mental models of sighted and visually impaired participants (or, later on, individualized combinations based on the preferences). To approach the question on "how to manipulate" the artifacts by PVIs, it is important to provide provision and also to research the integration of a) established and b) innovative modalities like head tracking or tangible objects. This research will yield the interaction concepts for the meeting room of the future as an "Accessible Meeting Room".

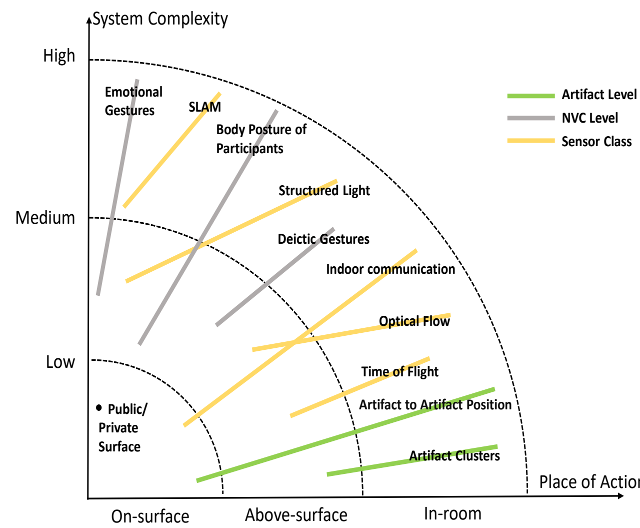

Figure 3: Sensor classes, NVC levels, and artifact levels

These concepts will be developed in a user-centered design approach and evaluated with PVIs together with sighted persons in mixed teams. The following sections will discuss this approach, in reference to Figure 3, in more detail.

Using multiple interactive surfaces shape a three-dimensional information space, which allows for various kinds of interaction and information distribution: on a surface, above a surface, and in the room. Regarding user interaction, some interactions become more complex to track, like e.g. pointing gestures in mid-air [34], while other gestures most likely occur with the distributed information space, such as e.g. deictic gestures, postures, and other NVC. While gestures and information distribution on surfaces are easy to detect, the system's complexity will increase with the size of the information space and the actions that can be performed within. Further, the captured information needs to be encoded into auditive and haptic actuations to make them accessible for PVIs.

How to capture?

Following the rules of a mediated Metaplan session, there are mainly two regions of interest to be captured by the sensors: (i) in front of the interactive vertical whiteboards, and (ii) on and around the (interactive) meeting table (see Figure 2). In both regions, video and audio need to be recorded. Since the video will only be used for tracking and to detect NVC elements, but not for video conferencing, any optical distortion can be tolerated as long as the relationship between objects is not affected. This allows using hemispheric lenses that could capture a large area when being placed on/above the table facing upwards/downwards (see Figure 1), or on top of the interactive whiteboards facing into the room. The cameras are installed in such a way that they capture widely overlapping regions. This allows capturing user actions from different perspectives, giving at least two images that can be correlated with each other to extract NVC elements in a reliable way.

Together with the video, an audio signal is captured and fed to the speech and voice recognition program that will allow identifying the active speaking person, but also to detect typical words that are accompanying deictic gestures such as "there", "here", etc.

How can a PVI filter this large stream of information?

MAPVI will employ a semi-automatic user interface approach to access relevant and to distinguish important from less important information. Depending on user capabilities, preferences and based on situations machine-learning technology can be facilitated to narrow down potentially relevant information. Additionally, the users themselves can actively retrieve information which is of importance to them, avoiding users to only become receivers of a broad stream of information, but rather transform their role to being users of an interactive system that assists during communication and reacts to their needs in a semi-automatic fashion.

How to make sense?

It is crucial that false alerts to the PVIs will be avoided in order to achieve and keep a high user acceptance. However, the acquired sensor data might be prone to errors (e.g., visually by camouflaging or occluding effects or by a noisy environment). Thus, the incoming sensor signals will undergo several filters and reasoning stages to make sense in the alerts to the PVIs. Depending on the incoming signal, the signal processing techniques (e.g., FFT) can be applied in order to allow for a more reliable analysis in the succeeding stages (see Figure 4).

Image analysis based on deep learning approaches. Since the most relevant regions in the meeting room are captured by multiple cameras, any action will be visible from different viewpoints. Real-Time Analysis of the images will allow recognizing the most common gestures like nodding, shrugging, or pointing gestures more reliably. Deep learning approaches have been proven to perform very well for the image processing tasks such as image classification, segmentation, registration, [40], etc., which further motivates us to use these approaches for emotion and gesture detection. Deep learning techniques will allow the learning of the relationship

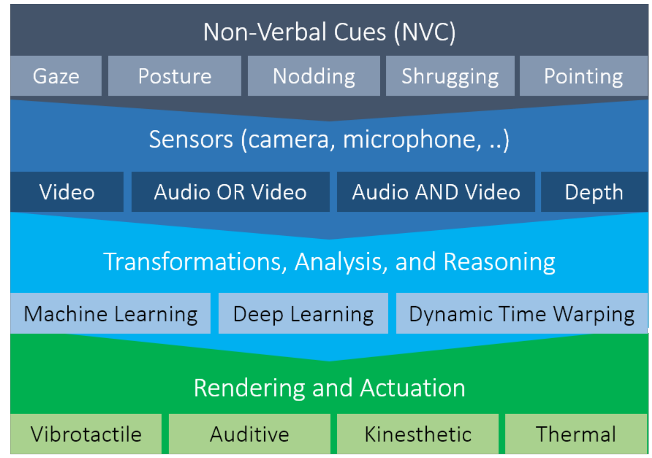

Figure 4: Signal processing pipeline with example technologies.

between the different entities by training on labeled or unlabelled data. This learning will help to get the real-time implementation which takes a few seconds to test in real time. The accuracy of the implementation can be improved by training on a large and clean dataset, which further requires us to create a dataset for training which has a similar meeting environment as used for real-time implementation.

Voice recognition. Deep learning approaches will furthermore be used for speech recognition, in particular, to detect keywords. For example, a set of keywords may describe a position in a three-dimensional information space or a second set may be used to identify deictic gestures. In addition, words describing features of an artifact (e.g. the color or the form of an object) will be detected, since they are also being used to describe positions. Moreover, an analysis of the formant frequencies in the frequency domain by deep learning will also allow identifying the speaker and his emotional state.

Temporal validity/synchronicity. Deictic gestures only make sense if they appear synchronously, i.e. a visible pointing gesture and a descriptive word. Whatever comes first (audio keyword or detected pointing gesture) will wait for a predefined time-span for the other input signal. If this additional signal does not appear, the first measurement will be discarded.

Reasoning. Reasoning will result in a refined measurement based on the fulfillment of the given premises. While each individual measure might be unsharp, logical reasoning gives more reliable results. For the given Metaplan session, a typical example could be the following case: A user says: "I mean the green card over there" and also perform the deictic pointing gesture towards a whiteboard. However, the speech detection is ambiguous and returns 'green' as 'gray'. Since the three-dimensional information space from Figure 2 offers an interactive table and two interactive whiteboards, and it might contain green and gray cards as well, it will not narrow down the results. However, there is the coarse pointing gesture towards one of the whiteboards that only contain green cards. This does not only resolve the ambiguity of the audio signal but also refines the precision of the pointing gesture, since the position of the digital artifacts (the cards), as well as their features (the colors), are precisely defined. Such multiple reasoning rules between the incoming sensor signals will help to further reduce false alerts to the PVI.

MTM classification. Human movements are already successfully classified in the industry by the so-called "Methods-Time Measurement" (MTM) (for a review see [39]), which basically splits up any human movement into elementary motions, such as reaching, grasping, releasing, etc. A similar taxonomy will be developed for typical brainstorming sessions, which might contain basic elements like "pointing", "shrugging", "nodding", etc. (see Figure 1). This classification will help to further refine the measurement behavior of the users since some of the basic elements might exclude each other, e.g. "nodding" and "shrugging".

Creating an ontology representing meeting content and environment. Since there is a huge amount of data being generated from various sources, a model needs to be developed to represent how these junks of information are structured and which relations exist among them. This includes information about meeting artifacts as well as peoples' gestures, facial expressions, locations, as well as verbal and textual contributions, e.g., on a whiteboard. Structuring, hierarchizing and explicitly describing relations between these junks of information is not only crucial for data storage, but also for data retrieval, since this would allow for more complex queries to only get relevant pieces for being presented to the PVIs who are highly prone to cognitive overload due to a rather limited sensory bandwidth because of missing eyesight.

How to actuate?

One of the key challenges of the MAPVI project will be to cope with the extensive amount of data that are sensed and reasoned to be relevant for a PVI. Thus one of the main questions of the project will be: *How to present this information to the PVI?*

The solution cannot simply be just exposing all information to the PVI and translating them to haptic and auditory channels. Thus, one of our first research questions will be: how can we address both haptic and auditory channels with different kinds of information? There are different types of haptic feedback, e.g. tactile vibration, electronic muscle stimulation, ultrasonic, pneumatic, kinesthetical haptic feedback, or even thermal haptic feedback as depicted in Figure 4. Our goal is it to simultaneously combine these haptic channels and assess the suitability of PVIs to simultaneously perceive different haptic stimuli. Similarly, for auditory cues, we want to test if PVIs can perceive differently pitched auditory information simultaneously.

Once we figured out the limits of the haptic and auditory channels, we will add information from a real meeting and perform user-studies to assess how this information is perceived and accepted by PVIs.

5 RESEARCH QUESTIONS AND NEXT STEPS

To achieve our goals to create accessible meetings for PIVs, we define the following research challenges.

Cognitive Model

When working in a three-dimensional information space, PVIs also perceive 3D information, e.g., by the voice from other users. This collides with existing interfaces like Braille displays that only provide a 1D output, which might prevent the PVI from building his own three-dimensional cognitive map and thus from performing deictic gestures by himself. We thus want to research whether no such three-dimensional cognitive map already exists and how this will change when providing new interfaces for 3D output to the PVI.

Haptic Actuation

After we identified the non-verbal cues and gathered all critical information during meetings, we need to translate those into haptic actuations. Hence, we will research a multitude of haptic sensations which can be directly used to encode information into haptics. Further, we want to know 1) how that haptic output can be rendered, and 2) how we can combine those haptic actuations to create a multi-modal meeting environment.

Next Steps

The next steps will be to examine the requirements of meetings and group collaboration more closely. We are planning several preliminary studies in which we will seek to identify all non-verbal cues as well as observe the verbal and gestural communication of the participants. In the next step, we will categorize and label the observed cues so that we can proceed with the interpretation of the found properties using a number of approaches, such as machine learning. Further, through interview rounds with PIVs we will collect additional feedback of end users. At the same time, we will develop first prototypes that can translate the cues into haptic stimuli and evaluate the combination of different approaches in user studies.

6 SUMMARY

In this project, we contribute novel concepts to assist persons with visual impairments in meeting scenarios and group collaborations by 1) capturing and understanding the overall environment, 2) reasoning the actions and properties, and 3) providing a multitude of actuations. This will help them in communicating their ideas and to get actively involved in the meeting conversation similar to a sighted person.

7 ACKNOWLEDGEMENTS

This work was commonly funded by DFG, FWF, and SNF under No. 211500647.

REFERENCES

[1] Bassam Almasri, Islam Elkabani, and Rached Zantout. 2014. An Interactive Workspace for Helping the Visually Impaired Learn Linear Algebra. In *Computers Helping People with Special Needs*, Klaus Miesenberger, Deborah Fels, Dominique Archambault, Petr Peňáz, and Wolfgang Zagler (Eds.). Springer International Publishing, Cham, 572–579.

[2] Michael Argyle, Mark Cook, and Duncan Cramer. 1994. Gaze and Mutual Gaze. *British Journal of Psychiatry* 165, 6 (1994), 848–850. https://doi.org/10.1017/S0007125000073980

[3] Mauro Avila Soto and Markus Funk. 2018. Look, a guidance drone! Assessing the Social Acceptability of Companion Drones for Blind Travelers in Public Spaces. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, New York, NY, USA, 417–419.

[4] Mauro Avila Soto, Markus Funk, Matthias Hoppe, Robin Boldt, Katrin Wolf, and Niels Henze. 2017. Dronenavigator: Using leashed and free-floating quadcopters to navigate visually impaired travelers. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, New York, NY, USA, 300–304.

[5] Amartya Banerjee, Jesse Burstyn, Audrey Girouard, and Roel Vertegaal. 2011. Pointable: An In-air Pointing Technique to Manipulate Out-of-reach Targets on Tabletops. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces (ITS '11)*. ACM, New York, NY, USA, 11–20. https://doi.org/10.1145/2076354.2076357

[6] Andrey V. Bogdanov. 2008. Neuroinspired Architecture for Robust Classifier Fusion of Multisensor Imagery. *IEEE Transactions on Geoscience and Remote Sensing* 46, 5 (May 2008), 1467–1487. https://doi.org/10.1109/TGRS.2008.916214

[7] Nikolaos Bourbakis. 2008. Sensing Surrounding 3-D Space for Navigation of the Blind. *IEEE Engineering in Medicine and Biology Magazine* 27, 1 (Jan 2008), 49–55. https://doi.org/10.1109/MEMB.2007.901780

[8] Stephen Brewster and Lorna M. Brown. 2004. Tactons: Structured Tactile Messages for Non-visual Information Display. In *Proceedings of the Fifth Conference on Australasian User Interface - Volume 28 (AUIC '04)*. Australian Computer Society, Inc., Darlinghurst, Australia, Australia, 15–23. http://dl.acm.org/citation.cfm?id=976310.976313

[9] Anke Brock, Philippe Truillet, Bernard Oriola, and Christophe Jouffrais. 2014. Making Gestural Interaction Accessible to Visually Impaired People. In *Haptics: Neuroscience, Devices, Modeling, and Applications*, Malika Auvray and Christian Duriez (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 41–48.

[10] Anke M. Brock. 2013. Touch the Map!: Designing Interactive Maps for Visually Impaired People. *SIGACCESS Access. Comput.* 105 (Jan. 2013), 9–14. Issue 105. https://doi.org/10.1145/2444800.2444802

[11] Michael Brock and Per Ola Kristensson. 2013. Supporting Blind Navigation Using Depth Sensing and Sonification. In *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication (UbiComp '13 Adjunct)*. ACM, New York, NY, USA, 255–258. https://doi.org/10.1145/2494091.2494173

[12] Sebastian Büttner, Henrik Mucha, Markus Funk, Thomas Kosch, Mario Aehnelt, Sebastian Robert, and Carsten Röcker. 2017. The design space of augmented and virtual reality applications for assistive environments in manufacturing: a visual approach. In *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments*. ACM, Ney York, NY, USA, 433–440.

[13] Rafael A. Calvo and Sidney D'Mello. 2010. Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications. *IEEE Transactions on Affective Computing* 1, 1 (Jan 2010), 18–37. https://doi.org/10.1109/T-AFFC.2010.1

[14] Christian Cherek, Simon Voelker, Jan Thar, Rene Linden, Florian Busch, and Jan O. Borchers. 2015. PERCs Demo: Persistently Trackable Tangibles on Capacitive Multi-Touch Displays. In *ITS*. ACM, New York, NY, USA, 389–392.

[15] Bruno R. De Araùjo, Géry Casiez, and Joaquim A. Jorge. 2012. Mockup Builder: Direct 3D Modeling on and Above the Surface in a Continuous Interaction Space. In *Proceedings of Graphics Interface 2012 (GI '12)*. Canadian Information Processing Society, Toronto, Ont., Canada, Canada, 173–180. http://dl.acm.org/citation.cfm?id=2305276.2305305

[16] Alexandra Delazio, Ken Nakagaki, Scott E Hudson, Jill Fain Lehman, and Alanson P Sample. 2018. Force Jacket: Pneumatically-Actuated Jacket for Embodied Haptic Experiences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI'18*. ACM, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173894

[17] Alan R. Dennis, Joey F. George, Len M. Jessup, Jay F. Nunamaker, and Douglas R. Vogel. 1988. Information Technology to Support Electronic Meetings. *MIS Q.* 12, 4 (Dec. 1988), 591–624. https://doi.org/10.2307/249135

[18] Scott Elrod, Richard Bruce, Rich Gold, David Goldberg, Frank Halasz, William Janssen, David Lee, Kim McCall, Elin Pedersen, Ken Pier, John Tang, and Brent Welch. 1992. Liveboard: A Large Interactive Display Supporting Group Meetings, Presentations, and Remote Collaboration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '92)*. ACM, New York, NY, USA, 599–607. https://doi.org/10.1145/142750.143052

[19] Sebastian Günther, Sven Kratz, Daniel Avrahami, and Max Mühlhäuser. 2018. Exploring Audio, Visual, and Tactile Cues for Synchronous Remote Assistance. In *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference on - PETRA '18*. ACM Press, New York, New York, USA, 339–344. https://doi.org/10.1145/3197768.3201568

[20] Sebastian Günther, Florian Müller, Markus Funk, Jan Kirchner, Niloofar Dezfuli, and Max Mühlhäuser. 2018. TactileGlove: Assistive Spatial Guidance in 3D Space through Vibrotactile Navigation. In *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference on - PETRA '18*. ACM, New York, NY, USA, 273–280. https://doi.org/10.1145/3197768.3197785

[21] Sebastian Günther, Martin Schmitz, Florian Müller, Jan Riemann, and Max Mühlhäuser. 2017. BYO*: Utilizing 3D Printed Tangible Tools for Interaction on Interactive Surfaces. In *Proceedings of the 2017 ACM Workshop on Interacting with Smart Objects - SmartObject '17*. ACM, New York, NY, USA, 21–26. https://doi.org/10.1145/3038450.3038456

[22] Wilko Heuten, Niels Henze, Susanne Boll, and Martin Pielot. 2008. Tactile wayfinder: a non-visual support system for wayfinding. In *Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges*. ACM, New York, NY, USA, 172–181.

[23] Otmar Hilliges, Shahram Izadi, Andrew D. Wilson, Steve Hodges, Armando Garcia-Mendoza, and Andreas Butz. 2009. Interactions

in the Air: Adding Further Depth to Interactive Tabletops. In *Proceedings of the 22Nd Annual ACM Symposium on User Interface Software and Technology (UIST '09)*. ACM, New York, NY, USA, 139–148. https://doi.org/10.1145/1622176.1622203

[24] Victoria E. Hribar and Dianne T.V. Pawluk. 2011. A Tactile-thermal Display for Haptic Exploration of Virtual Paintings. In *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '11)*. ACM, New York, NY, USA, 221–222. https://doi.org/10.1145/2049536.2049577

[25] Hans-Christian Jetter, Harald Reiterer, and Florian Geyer. 2014. Blended Interaction: understanding natural human–computer interaction in post-WIMP interactive spaces. *Personal and Ubiquitous Computing* 18, 5 (01 Jun 2014), 1139–1158. https://doi.org/10.1007/s00779-013-0725-4

[26] Brad Johanson, Armando Fox, and Terry Winograd. 2002. The Interactive Workspaces project: experiences with ubiquitous computing rooms. *IEEE Pervasive Computing* 1, 2 (April 2002), 67–74. https://doi.org/10.1109/MPRV.2002.1012339

[27] Udo Kannengiesser and Stefan Oppl. 2015. Business Processes to Touch: Engaging Domain Experts in Process Modelling. In *13th International Conference on Business Process Management*. BPM, Innsbruck, Austria, 40–44.

[28] Alexey Karpov and Andrey Ronzhin. 2014. A Universal Assistive Technology with Multimodal Input and Multimedia Output Interfaces. In *Universal Access in Human-Computer Interaction. Design and Development Methods for Universal Access*, Constantine Stephanidis and Margherita Antona (Eds.). Springer International Publishing, Cham, 369–378.

[29] Adam Kendon. 2010. *Spacing and Orientation in Co-present Interaction.* Springer Berlin Heidelberg, Berlin, Heidelberg, 1–15. https://doi.org/10.1007/978-3-642-12397-9_1

[30] Florian Klompmaker, Karsten Nebe, and Julien Eschenlohr. 2013. Towards Multimodal 3D Tabletop Interaction Using Sensor Equipped Mobile Devices. In *Mobile Computing, Applications, and Services*, David Uhler, Khanjan Mehta, and Jennifer L. Wong (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 100–114.

[31] Marc L. Knapp, Judith A. Hall, and Terrence G. Horgan. 2013. *Nonverbal Communication in Human Interaction.* Wadsworth CENGAGE Learning, Boston, MA, USA.

[32] Markus Koppensteiner, Pia Stephan, and Johannes Paul Michael Jäschke. 2016. Moving speeches: Dominance, trustworthiness and competence in body motion. *Personality and Individual Differences* 94 (2016), 101 – 106. https://doi.org/10.1016/j.paid.2016.01.013

[33] Sreekar Krishna, Greg Little, John Black, and Sethuraman Panchanathan. 2005. A Wearable Face Recognition System for Individuals with Visual Impairments. In *Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility (Assets '05)*. ACM, New York, NY, USA, 106–113. https://doi.org/10.1145/1090785.1090806

[34] Andreas Kunz, Ali Alavi, and Philipp Sinn. 2014. Integrating pointing gesture detection for enhancing brainstorming meetings using kinect and pixelsense. In *Procedia CIRP - Proceedings of the 8th International Conference on Digital Enterprise Technology (DET2014)*, Vol. 25. Elsevier, Paris, France, 205–212.

[35] Andreas Kunz and Morten Fjeld. 2010. *From Table-System to Tabletop: Integrating Technology into Interactive Surfaces.* Number Tabletops - Horizontal Interactive Displays in Human-Computer Interaction. Springer Verlag, Berlin, Germany, 53–72. http://www.springer.com/computer/user+interfaces/book/978-1-84996-112-7

[36] Sri Hastuti Kurniawan and Alistair G. Sutcliffe. 2002. Mental Models of Blind Users in the Windows Environment. In *Proceedings of the 8th International Conference on Computers Helping People with Special Needs (ICCHP '02)*. Springer-Verlag, London, UK, UK, 568–574. http://dl.acm.org/citation.cfm?id=646269.684511

[37] Orly Lahav and David Mioduser. 2001. Multisensory Virtual Environment for Supporting Blind Persons' Acquisition of Spatial Cognitive Mapping – a Case Study. In *Proceedings of EdMedia + Innovate Learning 2001*, Craig Montgomerie and Jarmo Viteli (Eds.). Association for the Advancement of Computing in Education (AACE), Norfolk, VA USA, 1046–1051. https://www.learntechlib.org/p/8729

[38] Saadi Lahlou. 2009. *Designing User Friendly Augmented Work Environments* (1st ed.). Springer Publishing Company, Incorporated, London, UK.

[39] J Laring, M Forsman, R Kadefors, and R Örtengren. 2002. MTM-based ergonomic workload analysis. *International journal of Industrial ergonomics* 30, 3 (2002), 135–148.

[40] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature* 521, 7553 (2015), 436.

[41] Carsten Magerkurth and Thorsten Prante. 2001. *„Metaplan" für die Westentasche: Mobile Computerunterstützung für Kreativitätssitzungen.* Vieweg+Teubner Verlag, Wiesbaden, 163–171. https://doi.org/10.1007/978-3-322-80108-1_18

[42] Nicolai Marquardt, Ken Hinckley, and Saul Greenberg. 2012. Cross-device Interaction via Micro-mobility and F-formations. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology (UIST '12)*. ACM, New York, NY, USA, 13–22. https://doi.org/10.1145/2380116.2380121

[43] Troy McDaniel, Sreekar Krishna, Vineeth Balasubramanian, Dirk Colbry, and Sethuraman Panchanathan. 2008. Using a haptic belt to convey non-verbal communication cues during social interactions to individuals who are blind. In *2008 IEEE International Workshop on Haptic Audio visual Environments and Games*. IEEE, New York, NY, USA, 13–18. https://doi.org/10.1109/HAVE.2008.4685291

[44] Albert Mehrabian. 1969. Significance of posture and position in the communication of attitude and status relationships. *Psychological Bulletin* 7, 5 (May 1969), 359–372.

[45] MICOLE. 2006. MICOLE - Multimodal Collaboration Environment for Inclusion of Visually Impaired Children. http://micole.cs.uta.fi/index.html

[46] Florian Müller, Joshua McManus, Sebastian Günther, Martin Schmitz, Max Mühlhäuser, and Markus Funk. 2019. Mind the Tap: Assessing Foot-Taps for Interacting with Head-Mounted Displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* https://doi.org/10.1145/3290605.3300707

[47] Stephan Pölzer and Klaus Miesenberger. 2014. Presenting Non-verbal Communication to Blind Users in Brainstorming Sessions. In *Computers Helping People with Special Needs*, Klaus Miesenberger, Deborah Fels, Dominique Archambault, Petr Peňáz, and Wolfgang Zagler (Eds.). Springer International Publishing, Cham, 220–225.

[48] Stephan Pölzer and Klaus Miesenberger. 2014. A Tactile Presentation Method of Mind Maps in Co-located Meetings. In *Proceedings of the International Workshop on Tactile/Haptic User Interfaces for Tabletops and Tablets, held in conjunction with ACM ITS 2014*. ACM, New York, NY, USA.

[49] Halley Profita, Reem Albaghli, Leah Findlater, Paul Jaeger, and Shaun K Kane. 2016. The AT effect: how disability affects the perceived social acceptability of head-mounted display use. In *proceedings of the 2016 CHI conference on human factors in computing systems*. ACM, New York, NY, USA, 4884–4895.

[50] Markus Rader, Clemens Holzmann, Enrico Rukzio, and Julian Seifert. 2013. MobiZone: Personalized Interaction with Multiple Items on Interactive Surfaces. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia (MUM '13)*. ACM, New York, NY, USA, Article 8, 10 pages. https://doi.org/10.1145/2541831.2541836

[51] Stefan Radomski and Dirk Schnelle-Walka. 2013. Spatial audio with the W3C architecture for multimodal interfaces. In *Workshop on Speech in Mobile and Pervasive Environments*. ACM, New York, NY, USA, 1–5.

[52] Georg Regal, Elke Mattheiss, David Sellitsch, and Manfred Tscheligi. 2016. TalkingCards: Using Tactile NFC Cards for Accessible Brainstorming. In *Proceedings of the 7th Augmented Human International Conference 2016 (AH '16)*. ACM, New York, NY, USA, Article 18, 7 pages. https://doi.org/10.1145/2875194.2875240

[53] Martin Schmitz, Mohammadreza Khalilbeigi, Matthias Balwierz, Roman Lissermann, Max Mühlhäuser, and Jürgen Steimle. 2015. Capricate: A Fabrication Pipeline to Design and 3D Print Capacitive Touch Sensors for Interactive Objects. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. ACM, New York, NY, USA, 253–258. https://doi.org/10.1145/2807442.2807503

[54] Eberhard Schnelle. 1982. *Metaplan Gesprächstechnik : Kommunikationswerkzeug für die Gruppenarbeit.* Metaplan GmbH, Quickborn, Germany. https://books.google.ch/books?id=3odkGwAACAAJ

[55] Dirk Schnelle-Walka, Stefan Radomski, and Max Mühlhäuser. 2013. JVoiceXML as a modality component in the W3C multimodal architecture. *Journal on Multimodal User Interfaces* 7, 3 (01 Nov 2013), 183–194. https://doi.org/10.1007/s12193-013-0119-y

[56] Dirk Schnelle-Walka, Stefan Radomski, and Max Mühlhäuser. 2014. Multimodal Fusion and Fission within W3C Standards for Nonverbal Communication with Blind Persons. In *Computers Helping People with Special Needs*, Klaus Miesenberger, Deborah Fels, Dominique Archambault, Petr Peňáz, and Wolfgang Zagler (Eds.). Springer International Publishing, Cham, 209–213.

[57] Dirk Schnelle-Walka, Chiew Seng Sean Tan, Patrick Ostie, Stefan Radomski, Kris Luyten, Karin Coninx, and Max Mühlhäuser. 2016. Whom-I-Approach: A System that Provides Cues on Approachability of Bystanders for Blind Users. http://tubiblio.ulb.tu-darmstadt.de/98343/

[58] Roy Shilkrot, Jochen Huber, Wong Meng Ee, Pattie Maes, and Suranga Chandima Nanayakkara. 2015. FingerReader: A Wearable Device to Explore Printed Text on the Go. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 2363–2372. https://doi.org/10.1145/2702123.2702421

[59] Joel Snyder. 2005. Audio description: The visual made verbal. *International Congress Series* 1282 (2005), 935 – 939. https://doi.org/10.1016/j.ics.2005.05.215 Vision 2005.

[60] Aureli Soria-Frisch, Alejandro Riera, and Stephen Dunne. 2010. Fusion operators for multi-modal biometric authentication based on physiological signals. In *International Conference on Fuzzy Systems*. IEEE, New York, NY, USA, 1–7. https://doi.org/10.1109/FUZZY.2010.5584121

[61] Caleb Southern, James Clawson, Brian Frey, Gregory Abowd, and Mario Romero. 2012. An Evaluation of BrailleTouch: Mobile Touchscreen Text Entry for the Visually Impaired. In *Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services (MobileHCI '12)*. ACM, New York, NY, USA, 317–326. https://doi.org/10.1145/2371574.2371623

[62] Norbert A. Streitz, Jörg Geißler, Torsten Holmer, Shin'ichi Konomi, Christian Müller-Tomfelde, Wolfgang Reischl, Petra Rexroth, Peter Tandler, and Ralf Steinmetz. 1999. i-LAND: An Interactive Landscape for Creativity and Innovation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI'99*. ACM, Pittsburgh, PA, USA, 120–127.

[63] Peter Tandler. 2000. Architecture of BEACH: The Software Infrastructure for Roomware Environments. In *ACM CONFERENCE ON COMPUTERSUPPORTED COOPERATIVE WORK (CSCW'2000*. ACM, Philadelphia, PE, USA, 2–6.

[64] Zhihong Zeng, Maja Pantic, Glenn I. Roisman, and Thomas S. Huang. 2009. A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, 1 (Jan 2009), 39–58. https://doi.org/10.1109/TPAMI.2008.52