ulm university universität ulm

RTMI ' 15

Proceedings of the
7th Seminar on

Research Trends in Media Informatics

13th February 2015

Institute of Media Informatics
Ulm University

institute of media informatics

RTMI '15

Proceedings of the 7th Seminar on Research Trends in Media Informatics

Institute of Media Informatics
Ulm University

Villa Eberhardt, Ulm
13th February 2014

Foreword by the Editors

RTMI'15 is the seventh incarnation of the annual seminar *Research Trends in Media Informatics* hosted by the Institute of Media Informatics at Ulm University. The RTMI seminar series aims to motivate students to delve deeper into the vast and diverse research in the area of media informatics, human computer interaction, and ubiquitous computing. Participants chose one of many suggested topics highlighting directions and challenges pursued by active research in this field. In order to provide participants with insights into the academic publishing cycle, the seminar emulates the process of submitting a technical paper to an academic conference. Participants first prepared their papers on a selected topic, either in English or German. The submissions were then peer reviewed in terms of content, academic quality, and presentation by at least two other seminar participants and one editor. The revised camera ready versions of the contributions constitute these proceedings. The authors presented their papers at the RTMI conference in February 2015, which was held in the Hall of Knights at the Villa Eberhardt in Ulm.

This year's program focuses around three themes, which underline that media informatics research covers far more than graphical user interfaces. The papers in the *Sensing and Tracking* session discuss novel technologies for sensor fusion, depth cameras in the hospital context, and non-optical object recognition. The *Technology and Society* session discusses social implications of technology, interaction concepts for people with disabilities, and provides an overview of techniques for projection mapping. The final session focuses on *Wearable Computing* and provides insights of microinteractions for wearable computing, interaction techniques for near-eye displays, and non-contact actuation of matter in the field of Human-Computer Interaction.

The editors would like to thank all authors for their effort and the work put into each individual contribution.

Ulm, February 2015

Jan Gugenheimer, David Dobbelstein, Katrin Plaumann, Florian Geiselhart
Christian Winkler, Philipp Hock, and Enrico Rukzio

Conference Program

Multi-sensor fusion: applications in human-computer interaction

Davut Kültür
Marktplatz 39
89312 Günzburg
davut.kueltuer@uni-ulm.de

ABSTRACT
This paper will illustrate an overview of multi-sensor fusion as they the originally used in military areas or in modern nonmilitary fields like medicine or entertainment. Further the technical details with different levels of fusion will be shown and examples of applications in Human-Computer-Interaction context will be analyzed and assessed. Finally in the conclusion part a summary over the technology and a prediction to the future will be given.

INTRODUCTION
Sensors are devices, which detect events and provide a corresponding output. The use of sensors became popular and effective parallel to the technological development. For example it would not be possible to interact with modern devices like smartphones without sensors, because these devices use sensors like touchscreens to capture the users input and interprate the given data to give an output. The topic multi-sensor fusion handles with the common use of several sensor combinations to achieve inferences that are not feasible from each individual sensor separately and make a higher level of interpretation possible. With the use of multi-sensors new dimensions can be reached and also the accuracy can be increased. As shown by Hall and Llinas [4] multi-sensor fusion was in immediate past a huge development in military areas, e.g. automated target recognition or battlefield surveillance. But in recent years also non-military fields used more and more multi-sensor fusion like medicine, e.g. multi-sensor image fusion for tomographies. Further multi-sensor fusion become more and more available for average people as a result of smartphones and gaming consoles. So what are the advantages of the technology, especially in Human-Computer-Interaction context?

CONCEPTS
Luo et al. [9] reviewed in their work different theories and approaches of multisensor fusion and integration. Their multi-sensor fusion architecture based on the model of Luo&Kay [8], which will also be used in this paper to explain the fusion. However there are other fusion models like the process model of the Joint Directors of Laboratories shown by Hall and Llinas [4], but this paper will be based on the model of Luo&Kay [8].

7th Seminar on Research Trends in Media Informatics (RTMI '15). February 2015, Ulm University, Ulm, Germany.

Fusion Model
The architecture can be structured into three abstract levels: low level, medium level and high level. The process of multi-sensor fusion can be performed at these levels, which will be more concrete in following sections. This model comprises the stimation of incoming signals at low level. The classification of extracted features is part of the medium level and the decision making according to symbols and subdecisions takes place in high level. [8]

Low Level Fusion
The low level fusion contains signal and pixel level fusion as shown by Luo et al. [9]

At signal level the sensory data represents the output and need to be synchronized and adapted before the fusion process. Statistial estimation methods have been successfully used here for data fusion and can be grouped into nonrecursive and recursive methods. Nonrecursive ones like the weighted average method or the least square method are only used to merge redundant data. Recursive estimation methods like the Kalman filter or the extended Kalman filters can be applied to more fusion purposes.

Pixel level fusion is based on the image processing on the original pixel information. After getting the information from individual sensor this level generate a new composite image with better quality and more features providing a better interpretation of the scene. Used methods are band rationing fusion, the principal component analysis, the wavelet transform fusion and the combined use of them. [9]

Medium Level Fusion
Also called Feature Level Fusion, this level fuses features extracted from signals and images. Feature points obtained from different sources are concatenated to result in a feature with higher discrimination. Generally this level can be decomposed into three steps: Feature set uniformization and normalization, feature reduction and concatenation, and feature matching. Methods like support vector machines, cluster analysis, k-means clustering, Kohonen feature map and learing vector quantization are frequently used at this level. [9]

High Level Fusion

This level is also called Symbol Level Fusion. The information here is the symbolic representation of process parameters. It refers to the combination of symbols with an assoicated uncertainty measure into a composite decision. Symbol fusion is also referred to decision fusion. Suitable for this level are algorithms which are tolerant of imprecision, uncertainty, partial truth and approximation. Also neural networks, genetic algorithms, evolution algorithms, and fuzzy logic are employed. Finally inference methods like Bayesian inference or Dempster-Shafer method are successfully applied herefor. [9]

Algorithms and Methods

This section will handle the most frequent algorithms implemented in multi-sensor fusion. They are defined by the information process and mainly stem from prohability theories, data classification methods, and articial intelligences. The main concepts of the Estimation Methods, Classification Methods, Inference Methods and Artificial Intelligence Methods will be introduced following.

Estimation Methods

The Kalman filter is a useful estimation method for engineering applications like computer vision or target tracking. For example we have a signal like a sound and want to discard the noise in the environment. With the Kalman filter it is possible, because it is an optimal estimator, which infers parameters of interest from indirect, inaccurate and uncertain observations. [6] An optimal estimation in a statistical sense can be provided if the system can be described as a linear model and the system and sensor error can be modeled as white Gaussian noises. The Kalman filter has the advantage of his computational efficiency due to the use of efficient matrix operations for estimation. The restriction of having a linear model and a initial Gaussian uncertainty can be solved with extended Kalman filters which linearize the system using Taylor series expansions around a stable operating point. Another challenge is the problematic use in distributed tracking fusion caused by inconsistency of different sensory sources. [9]

Other estimation methods like the covariance intersection or covariance union are covariance-based fusion algorithms. The covariance intersection solves the fusion problem of the Kalman filter that it is awkward. If the sensory data measured from multiple sensors are not independent, it can yield a consistent estimate which is independent of the correlation between sensory data. The covariance union algortihm solves in addition the problem of information corruption. Herefor inconsistent estimates are replaced with a single estimate which is statistically consistent with all given estimates. The estimate is consistent as long as at least one estimate is consistent. [9]

Classification Methods

The classification methods are similar to the processes grouping data from multiple sources into classified datasets. A multidimensional feature space is first partitioned into distinct classes and the place of the new coming feature vector is compared with preclassified locations in the feature space. So that one can identify which data class the new feature attrbitues. There are two groups of classification techniques, parametric and nonmparametric ones. The parametric methods include clustering approaches, successfully applied in data mining, machine learning and pattern recognition and parametric templates, widely applied in image processing and computer vision. [9]

Unlike the parametric methods, the nonparametric classification algorithms like the Support Vector Machine, first proposed by Vapnik et al. [1], are not constrained to prior assumptions on the distribution of input data. The task of these algroithms is detecting and exploiting complex patterns in data e.g. by clustering classifying or ranking. Typical problems are hor to represent complex patterns and how to exclude spurious patterns. [2] The Support Vector Machine generates an optimized discriminant, called hyperplane, to demarcate the training data into two classes. On an optimal hyperplane the classification possesses minimum errors and the maximum margin between these classes.

Inference Methods

There exist two popular inference methods, Bayesian inference and D-S reasoning. the first one can address most of the fusion problems more efficiently than the second one. In contrast the D-S reasoning makes explizit any lack of information concerning a propositions probability and can address more problems. [9]

A Bayesian filter is a probabilistic estimation method. It uses for that a recursive predict-update process. The (extended) Kalman filter and the particle filter are both Bayesian-type algorithms, which are frequently adopted. The particle filter is based on point mass representation of posterior probability density. The main advantage is the ability to represendt arbitrary probability densities in nonlinear and non-Gaussian systems. After Luo et al. [9], the idea behind the particle filter is to represent the reqiured posterior density function by a set of random samples with associated weights and to compute estimates based on these samples and weights. Particle filters are often used in the application of mobile robots and tracking of targets.

The second inference method is the D-S evidence theory and is based on two ideas. First is obtaining degrees of belief for one question from subjective probabilities for related questions and the second is using Dempsters rule to combine the degrees of belief when they are based on items of evidence independently. This technique is usually implemented with different algorithms to improve the accuracy of the decision. For example Zhan et al. [x] combined this theory with the genetic algorithm to improve the accuracy of the gender and age recognition in a home service robot. The results showed that the accuracy of recognition was better than before. [9]

Artificial Intelligence Methods

Artificial intelligence methods are used at high level fusion to make a joint decision from local decisions of multiple sensors and can be seen as an advanced version of the estimation, the classification and the inference method at low level fusion. To fit complex nonlinear relationships with necessary precautions, they can be model free and have also sufficient degree of freedom. [9]

The artificial neural network is such a method for high level inference. It consists of layers of processing elements which may be interconnected in variety of ways. Neurons can be trained to represent sensory information. Also complex combinations of these can be activated in response to different sensory stimuli through associative recall. The extracted features from multiple sensors are classified by the artificial neural network and are merged to yield more accurate results. The incorporation of different algorithms improves also the system performances shown by Mitra et al. [10].

APPLICATION EXAMPLES

Your Phone or Mine? Fusing Body, Touch and Device Sensing for Multi-User Device-Display Interaction

Rofouei et al. [11] presented their technique ShakeID for associating multi-touch interactions to individual users and their mobile devices with the accelerometer sensors to determine who is interacting with a multi-user touch display. Therefore the real-time accelerometer data and the depht camera-based body tracking are used to associate each phone with specific user and also body tracking and touch positions are compared to associate a touch contact with the particular user. Their target was to make an association of touch contacts with devices possible to allow more seamler device-display multi-user interactions.

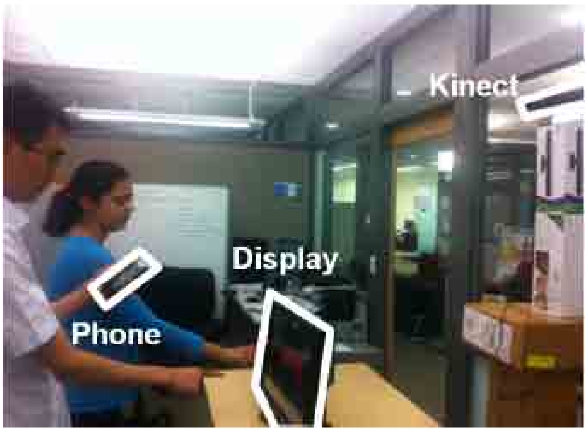


Figure 1. System with Kinect camera, multi-touch display and 2 accelerometer phones.

[11]

ShakeID uses multi-sensor fusion with the combination of the on-board sensors of the mobile devices and the touch sensing to perform an association. Additionally it matches the motion sensed by the device to motion observed by a Miscrosoft Kinect camera pointed at the users position in front of the touch screen. The system can associate each phone to a specific users hand by comparing the motion of each phone in scene with the motion of each user. Finally the performation of a coordinate transform from the 2D space of the screen to the 3D camera space causes the association of the tpuches on the display to users.

In detail ShakeID uses two steps. First to associate personal private smartphones to users holding them and second to associate touches on the shared screen to users who performed those touches. Rofouei et al. [11] assumed in their work that the smartphones had been previously paired to the system and focus on identifying the device. Their technique was implemented using the Microsoft Kinect for Windows SDK to track the hands of multiple users, the Miscrosoft Surface 2.0 SDK for the multi-touch display and two Winows Phone smartphones. [11]

The algorithm they used relies on the fact that if users hold their phones in their hands the acceleration measured by the phone accelerometers should match the acceleration of the hands holding the phones. They continuously correlated phone acceleration for each phone connected to the system with the accelerations of all hands tracked by the Kinect to associate the phone with a users hand. The 3-axis accelerometer in the phone captured some data which was sent wirelessly to the siplay system continuously. To estimate acceleration of hand position over time Rofouei et al. [11] used a Kalman filter, described before, with his advantage of good computational efficiency. Furthermore their algorithm associated touch contacts on the display to users hand positions. They first converted the 2D display coordinates of each contact to Kinects 3D coordinate system by a linear coordinate transform determined in an offline calibration. The hand corresponding to each contact is found by comparing 3D hand positions to transformed contact coordinates and after that each touch contact was mapped to a phone by the system.

The ShakeID technique has also limitations. In the description before one important limitation involves the case where the hand holding the phone is stationary. Rofouei et al. [11] agreed that in this case the matching process is likely to match equally well to other hands and in practical scenarios it is less likely that people will be stationary. Another limitation was the distance to the Kinect sensor. This should be not too small to compute correctly skeleton data.

A user study was also created to evaluate the ShakeID, where a system for content sharing between phones and devices using ShakeID was implemented. Seven pairs of participants were conducted and each study consisted of two people performing a set of content sharing tasks.

The main user study included two parts, the parallel and collaborative use. In parallel use part, the participants worked side-by-side and conducted 20 copies in each side trained before for every participant. The other part aimed to simulate collaborative actions where participants shared and discussed content through a shared display. The participants copied shapes between them therefore. The results was that during the parallel task 94% and during the collaborative task 92% accuracy was observed. The errors occurred when the hand holding the phone moved out of the field-of-view of the Kinect and limited the availability of accurate position data. [11]

Providing the Basis for Human-Robot-Interaction: A Multi-Modal Attention System for a Mobile Robot

Lang et al. [7] developed a robot which uses multi-modal data fusion to interact naturally with humans around his environment using a pan-tilt camera for face recognition, two microphones for sound source localization and a laser range finder for leg detection. In their paper, they showed an attention system for a mobile robot which enables the robot to shift its attention to the person of interest and to maintain attention during interaction.

In order to enable a natural interaction with robots, they have to be able to recognize automatically when a persons attention is directed towards it for communication. Therefore the detection of the communication partner is very important. Lang et al. [7] used a method for multi-modal person tracking which uses sensor data from a camera, two microphones and a laser. After fusing this data and making a decision, the robot can e.g. turn the direction of camera into the direction of a person to shift his attention and make the person to a person of interest.

The mobile robot for this purpose was called BIRON and was a pioneer PeopleBot from ActivMedia with an on-board PC to control the motors and the on-board sensors and to process sound. Another PC inside the robot was used for image processing. Two AKG far-field microphones used to communicate handfreely and a SICK laser range finder was mounted at the front to detect legs of humans.

The key part of their method was the multi-modal person tracking as a base of following processes. To distinguish between different persons it is necessary for a robot like BIRON to track all persons present robustly. Possibly the persons and the robot will be moving constantly so the sensory perception is always changing. A complex object like a person can not be captured by asingle sensor system alone, so BIRON obtains different percepts of a person.

The camera was used to recognize faces and has two steps. First the distance, direction and height of an observed person was extracted and then the identity of the person was given if it is known to the system. Two stereo microphones were applied to locate sound sources using a method like a Cross-Powerspectrum Phase Analysis [3] and made an estimation of the direction relative to the robot possible. The laser range finder was used to detect the legs of persons, which could be easily done because human legs are resulting in a characteristic pattern in range readings. From this data the distance and the direction of the person relative to the robot could be extracted.

In order to fuse this information they used a multi-modal-framework called multi-modal anchoring. The goal was to establish connections between processes that work on level of abstract representations of objects in the world, which could be defined as the symbolic level and processes that are responsible for the physical observation of these objects, the sensory level. These connections called anchors and must be dynamic, since the same symbol must be connected to new percepts every time a new observation of the csorresponding object is acquired. Multi-modal anchoring allows to link the symbolic description of a complex object to different types of percepts from different perceptual systems. Distributed anchoring of individual percepts from multiple modalities and copes with different spatio-temporal properties of the individual percepts are enabled. Every part of the object is captured by one sensor and this sensor is anchored by a single component anchoring process. A composite anchoring process realizes the composition of all component anchors and establishes the connection between the symbolic description of the complex object and the percepts from individual sensors. The composite anchoring model requieres a composition model, a motion model and a fusion model.

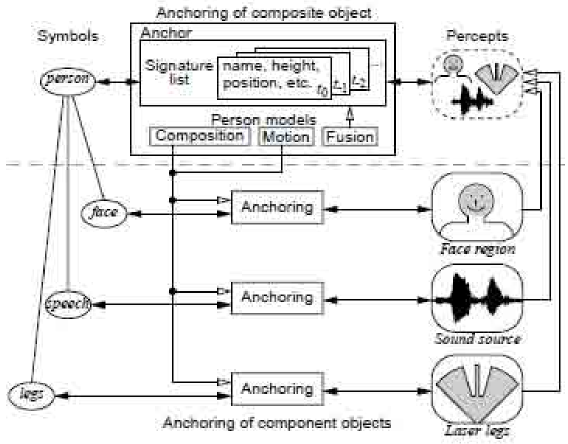


Figure 2. Multi-Modal Anchoring.
[7]

The composition model defines the spatial relationships of the components with respect to the composite object and the motion model describes the type of the motion of the complex object and allows to predict its position. The composition model is used in the component anchoring process to anchor only those percepts that satisfy the composition model. The information in the motion model can be used by the component anchoring processes in different ways. The component anchoring process selects the percept which is closest to the predicted position if multiple percepts are generated from one perceptual system. Or the perceptual system turns the sensor into the direction of the predicted person if the corresponding perceptual system receivees its data from a steerable sensor. The fusion model defines how the perceptual data from the component anchors has to be combined. The perceptual data may not arrive at the composite anchoring process in chronological order, because the processin times of the differen perceptual systems may differ significally. The composite anchor provides a chronologically sorted list of the fused perceptual data, inserts data from the component anchors in the list and updates all subsequent entries.

In the case of tracking multiple persons, multi-modal anchoring may lead to different conflict between the individual composite anchoring processes. The anchoring processes may try to control the pan-tilt unit of the camera in a contradictory way or a percetp may be selected by more than one anchoring process. Lang et al. [7] used a supervising module to solve the problems. it restricted the acces to the pan-tilt unit of the camera only one composite anchoring process at a time and solved the first problem. The second problem was also solved as follows. Every component anchoring process assigned scores to all percepts rating the proximility to the predicted position and the supervising module computed the optimal non-contradictory assignment of percepts to component anchors.

In summary Lang et al. [7] created for their paper a robot which was able to collect sensory information from different sensors to track complex objects like persons and applied multi-modal-anchoring to handle the data. A composite anchoring process realized the composition of all component anchors and requierd a composition model, motion model and a fusion model.

Monitoring Intake Gestures using Sensor Fusion (Microsoft Kinect and Inertial Sensors) for Smart Home Tele-Rehab Setting

Hondori et al. [5] had the goal to help post-stroke patients by developing smart home technology which supports these patients to complete activities of daily living independently. Beside the should save their time, money and extra effort. They presented in their paper an approach to spot specific activities of daily living of eating and drinking in a home setting. They fused therefore inertial and Microsoft Kinect sensors to monitor the patients intake gestures including fine cuttinm loading foot and manuevering the food to the mouth. They measured for both sides of the body first, the position of the wrist, elbow and shoulder, second, angular displacements at the elbow and shoulder joints and third, the acceleration of the spoon, far or cup which are held by the subject. The use of the Kinect sensor allowed them to distinguish between healthy and paralyzed body sides which is a common problem in tele-rehab. [5]

The purpose of their work was to improve the accuracy and efficacy of the therapy of post-stroke patients by using a tele-rehabilitation technology within patients homes. Since it is recommended to continue rehabilitation until maximxum recovery has been achieved and due to cost factors of outpatient rehabilitation facilities, the extension of chronic stroke-patientscare is limited, the tele-rehabilitation seems to be a good alternative. Smart houses could solve the problems described before, including devices to monitor the patientshealth status by helping them function more independently and also receive feedback on their activities.

They decided to use intake gestures like eating and drinking activities, which are critical functional activities for recovery and are among commonly trained activities during conventional therapy sessions. The movements require upper body movements including arm and trunk. They characterized these movements into food and beverage consumption such as holding fork, spoon, knife or cup, loading the food and actial food intake. Their setup included inertial sensors, which are MEMS devices that measure and report on acceleration and the Microsoft Kinect sensors, known for providing full-body 3D motion capture, facial recognition and voice recognition capabilities. The inertial sensors are attached to each of these eating utensils which record their local movements mady by the subject. The patient was also be observed by the Kinect sensor which has been placed on the same table while eating.

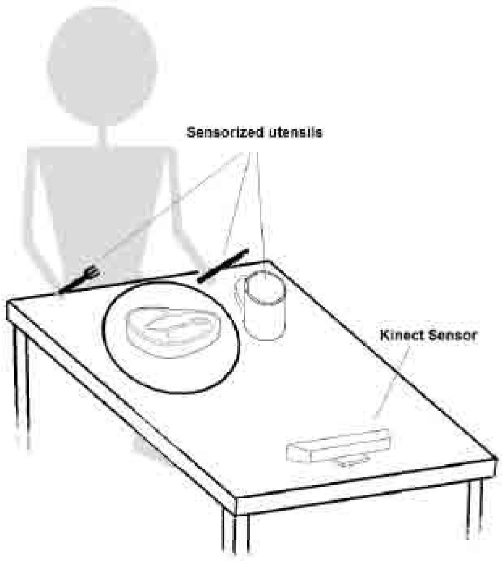


Figure 3. The system setup including the dining table, sensorized tools and a Kinect sensor.

[5]

Experimental data was also collected by the subject, which performed several movements simulating eating soup with his right hand at comfortable speed and then repeated the task of cutting a piece of stake with his right hand and maneuvering it to this mouth with his left hand for a couple of times. Finally he was instructed to perform drinking water by picking the cup with his right hand, holding it still at the mouth for a few seconds and repeating the task again for several times. With the Kinect sensor Hondori et al. [5] could analiyze the changes in position of different upperlimb joints in a body skeleton including the subjects head, center shoulder, right or left shoulders of elbows and wrists while performing the previous tasks. The wrists and elbows of the patient moved the most among the joints confirming the large maneuvering movements between the food plate on the table and his mouth wihle his center shoulder and head were still. Also the change of angles between different upper-limb joints of the subject while eating and drinking could be seen. During the manuervering movement, the right-elbow angle changed between approximately between 50 to 100 degrees and the left-elbow anlge varies between approximately between 90 to 110 degrees. The analyze of the inertial sensor data demonstrated that due to gravitational accelerations, the signal of the sensors alwas carried a bias equal to 9.81 m/s^2. By determining the direction of gravity, this bias was removed from the inertial sensors of each measurement unit. In case of cutting the stake, frequency of the movements had the highes value, whereas while moving the food to the mouth, frequency and magnitude of the movement was steadier. [5]

Hondori et al. [5] developed in their paper a low-cost, home-based system that is able to monitor patientsspecific activity of daily livings of eating and drinking to asses their level of progress over time. Their sensor fusion system leveraged inertial and Microsoft Kinect sensors to monitor arm gestures related to typical meal intake. They demonstrated that position, angular displacement and acceleration of arm gestures can be captured and analyzed. The conclusion was that sensor fusion can achieve more accurate monitoring performance as it uses the information received by two sources rather than one. [5]

CONCLUSION

This paper gives an introduction to multi-sensor fusion by describing the concepts based on the model of Luo&Kay [8]. The fusion architecture can be structured into three levels, in which the low level fusion contains signal and pixel level fusion. Signal level represents the sensory data as the output which needs to be synchronized and adapted before the fusion and pixel level fusion is based on the image processing which generates a new composite image with better quality and provides a better interpretation of the scene. In Medium level fusion the features extracted from signals and images are fused. The third level is the high level fusion, also called sybol level fusion and refers to the combination of symbols with an assoicated uncertainty measure into a composite decision.

Furthermore Algorithms and Methods frequently implemented in multi-sensor fusion are described. These algorithms can be grouped into estimation methods, classification methods, inference methods and artificial methods. Estimation methods include the Kalman filter with the advantage of his computational efficiency and other estimation methods like the covariance-based algorithms, which solves the awkwardness-problem of the kalman filter. Classification methods include parametric and nonparametric algorithms, in which parametric methods contain clustering approaches, machine learning and pattern recognition. The nonparametric ones like the Support Vector Machineare not constrained to prior assumptions on the distribution of input data. Inference methods include two popular algorithms, Bayesian inference and D-S reasoning. The Bayesian inference is a prohabilistic estimation method which uses a recursive predict-update process. D-S reasoning is based on the idea obtaining degrees of belief for one question and using Dempsters rule to combine the degrees of belief. Lastly artificial intelligence methods like the artificial neual network are used at high level fusion to make a joint decision from local desicions of multiple sensors.

The last part contains three different examples of applications, which use multi-sensor data fusion to gain better results in their works. Rofouei et al. [11] presented their technique ShakeID for associating multi-touch interactions to individual users and their mobile devices with the accelerometer sensors to determine who is interacting with a multi-user touch display. It used multi-sensor fusion with the combination of the on-board sensors of the mobile devices and the touch sensing to perform an association. The second example is from Lang et al. [7], they developed a robot which uses multi-modal data fusion to interact naturally with humans around his environment using a pan-tilt camera for face recognition, two microphones for sound source localization and a laser range finder for leg detection. In order to fuse the information they collected from the sensors they used a multi-modal-framework, which is also described in detail. The last example is from Hondori et al. [5], which developed in their paper a low-cost, home-based system that is able to monitor patientsspecific activity of daily livings of eating and drinking to asses their level of progress over time. Their sensor fusion system leveraged inertial and Microsoft Kinect sensors to monitor arm gestures related to typical meal intake.

REFERENCES

1. Cortes, C., and Vapnik, V. Support-vector networks. *Machine Learning 20*, 3 (Sept. 1995), 273–297.
2. Cristianini, N. Support Vector and Kernel Machines.
3. Giuliani, D., Omologo, M., and Svaizer, P. Talker localization and speech recognition using a microphone array and a cross-powerspectrum phase analysis. In *In Int. Conf. on Spoken Language Processing (ICSLP* (1994), 1243–1246.
4. Hall, D. L., Member, S., and Llinas, J. An Introduction to Multisensor Data Fusion.
5. Hondori, H. M., Khademi, M., and Lopes, C. V. and Inertial Sensors) for Smart Home Tele-Rehab Setting.
6. Kleeman, L. Understanding and Applying Kalman Filtering.
7. Lang, S., Kleinehagenbrock, M., Hohenner, S., Fritsch, J., Fink, G. A., and Sagerer, G. Providing the Basis for Human-Robot-Interaction : A Multi-Modal Attention System for a Mobile Robot. 28–35.
8. Luo, R., and Kay, M. A tutorial on multisensor integration and fusion. *[Proceedings] IECON '90: 16th Annual Conference of IEEE Industrial Electronics Society* (1990), 707–722.
9. Luo, R. C., and Chang, C.-C. Multisensor Fusion and Integration: A Review on Approaches and Its Applications in Mechatronics. *IEEE Transactions on Industrial Informatics 8*, 1 (Feb. 2012), 49–60.
10. Mitra, V., Wang, C.-J., and Banerjee, S. Lidar detection of underwater objects using a neuro-SVM-based architecture. *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council 17*, 3 (May 2006), 717–31.
11. Rofouei, M., Wilson, A., a.J. Bernheim Brush, and Tansley, S. Your Phone or Mine? Fusing Body, Touch and Device Sensing for Multi-User Device-Display Interaction. *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12* (2012), 1915–1918.

Depth Cameras in Health Care and Hospital Contexts

Katerina Tzikas
Media Computer Science
University of Ulm
katerina.tzikas@uni-ulm.de

ABSTRACT
Depth cameras provide a unique tool in health care that can help track patient rehabilitation or function as an easy-to-use input device for in-home monitoring systems and even in the ICU. In-home rehabilitation is a hot research topic in health care with the intent of reducing overall costs, making space in care facilities and freeing up time of doctors and therapists. Depth camera based body tracking helps to detect and even predict falls of patient's in a non-intrusive way.
This paper gives an introduction to a variety of applications of depth cameras in current research. Different techniques are going to be shown and compared to each other, and the overall up- and downsides of depth cameras in the context of health care are going to be presented.

Author Keywords
Kinect; Depth camera; Health care; Gait measurement; Monitoring; Telerehabilitation; Touchless Userinterface

ACM Classification Keywords
H.5.2. Information Interfaces and Presentation: User Interfaces; I.1.2.10. Artificial Intelligence: Vision and Scene Understanding; I.4.8. Image Processing and Computer Vision: Scene Analysis

INTRODUCTION
There are a number of fields in health care that may benefit from the use of body tracking devices. Be it as touchless user interface in surgical procedures, as in-home monitoring system in rehabilitation or as gait measurement tool in diagnosis. Especially human body gait is an important health indicator in all kinds of areas such as diabetes, neurological diseases and fall detection and prediction of stroke patient's and the elderly.
Known methods for 3D body tracking include marker based systems (typically making use of infrared markers), wearable sensors like force plate sensors, visual cameras and last but not least depth cameras.

7th Seminar on Research Trends in Media Informatics (RTMI '15). February 2015, Ulm University, Ulm, Germany.

Depth cameras, as their name suggests, do not gather RGB information like colour cameras do, but instead capture depth information. The resulting information yields an abstract representation of the scene. This is why depth cameras are considered non-intrusive when monitoring, contrary to visual cameras. Additionaly they do not need any additional devices to be placed on the user, as opposed to marker based systems and wearable sensors.
These properties make depth cameras especially applicable in health-care, where the patient's privacy is to be considered as well as the patient's ability (or disability!) to use a certain device by himself.
In the following I will shortly explain the fundamentals of operation of depth cameras, then elaborate on a few research projects in which depth cameras have been used.

DEPTH CAMERA BASICS
There are two different approaches to obtaining depth information of a scene. In general two basic components are needed: an emitter (a projector) and a sensor. One of these approaches makes use of a laser projector and a corresponding sensor to deduce depth information by measuring time-of-flight (TOF) of projected light.
Cameras operating with TOF sensors, like the Mesa Imaging TOF Camera, have a very high price range, beginning with several thousands of Euros. The second approach to obtaining depth information is based on triangulation of an infrared (IR) pattern emitted and detected by an IR-projector and an IR-sensor. The computed depth points result in a depth map that can further be analysed.
This approach is used by the Microsoft Kinect [10], as

Figure 1. Camera components of the Kinect
[10]

shown in Figure 1. The Kinect has been released in 2010, marking a low cost breakthrough for depth cameras, and has since been used in numerous research projects. Furthermore an Software Developement Kit (SDK) for the Kinect has been available since 2011, including a method for the so-called skeletal tracking. Skeletal tracking uses the depth points of a persons body and classifys each one of them into different body parts. Up to 20 body joints can be deduced from the depth image of a human. These joints are essential in interpreting a persons body posture and movement.

Figure 2. Original Picture

Figure 3. Depth image: Lighter areas are closer to the camera, while black pixels indicate missing depth values.

At a framerate of 30fps the Kinect grants an efficient and precise measurement.

Figure 4. Skeletal Tracking
[10]

Alternative tracking devices are visual devices like RGB cameras and sensoral devices like pressure sensors, accelerometers and wearable IR sensors.

RESEARCH AREAS

Research areas for the application of depth cameras include in-home rehabilitation and monitoring, touchless user interface design, gait analysis and sleep cycle detection. I will expand on these applications in the following sections.

Touchless user interface in surgical procedures

In the operating room it is essential for everything to be cleansed and sterile. Nothing unnessecary to the procedure has to be be lying around in order to provide an easy-to-use and efficient environment for the surgeon. Easy-to-use surgical setups reduce the risk of errors that may lead to life threatening complications. A touchless user interface for applications such as visualisation software is an ideal solution to keep everything clean and simple.
To this end, Ruppert et al. developed and evaluated a gesture-to-mouse mapping interface using the Kinect 1.0 depth camera [5]. Their research was done prior to the Kinect SDK release. For their two prototypes they used the then available libraries Libfreenect and OpenNI in combination with NITE. These libraries were nessecary to access the Kinect depth information. Only OpenNI/NITE included body recognition algortithms.

Their first prototype using Libfreenect needs to have the distance of the user's hand to the sensor predefined, in order to to be able to distinguish the hand from the rest of the body. The definition of this threshold proves to be unstable and noisy, thus noise reducing filters have to be applied on top of the threshold.
Once the points of interest (i.e. the depth values of the hand) have been determined, the center of mass (COM) of the point cloud is calculated. The COM serves as defining reference point that is then mapped to the mouse input of the operational software. A left mouse click is interpreted only if the hand is being hold still for one second.
The downsides of this first prototype include the possibility of involuntary mouse clicks and the restriction of the preset distance between hand and tracking device.

The second prototype improves on these downsides. With OpenNI/Nite being able to interpret certain body joints, no predefined distance is needed to distinguish the hand. The position of the hand can instead be deduced from the position of the user's torso joint. This means that the user can be standing anywhere in the operating room to operate the software. Unintentional mouse clicks are reduced by using the other hands height in relation to the torso joint as mouse click input. This feature furthermore allows the user to perform drag events. Ruppert et al. use a specic visualisation software for the monitoring of tumors during operations, but the method of input mapping to the mouse pointer is so versatile that it can be applied to any program.

In their evaluation they conlude that the interface is not only easy to learn, but also grants an efcient, sterile and non-invasive tool during operations with less distractions and therefor lower risk of erros than conventional keyboard and mouse input methods provide.

Rosa and Elizindo argue in [4] that in this setup the user has to step out of a wide interaction zone to prevent unwanted input and that he might be fatigued from the nessecary wide arm movements. Instead they suggest a method using a short ranged, high resolution novel motion sensor called Leap Motion sensor. Contrary to the Kinect this sensor accumulates data from an upward facing cone-like area, using multiple IR projectors and sensors and determining the 3D positions with undisclosed algorithms.
In [3] Jacob et al. seek another approach to preventing unwanted input while still using the Kinect. They train a decision tree with a collection of data, in order to define the user's intent by contextual cues. With this method they manage to reduce False-Positive inputs to an astonishing minimum of just over 1%.
Other approaches to touchless interfaces that are not based on depth camera input usually require the user to be wearing sensoring devices on the wrist or under the surgeon glove and therefor lack convenience.

Telerehabilitation

An examplatory application of depth cameras in telerehabilitation can be seen in [1]. The Kinect Rehabilitation System (KiReS) is a software that uses the Kinect to create, perform and track in-home rehabilitation excersises.

The software inlcudes two user frontends: One for the therapist to create exercise sessions, as well as to review collected data, and one for the patient to perform these exercises. The tasks are shown to the patient by the use of two avatars. One of them representing the exact exercise to be performed, the other one visualising the user's currently tracked body position and movement. The user can easily compare the two avatars and asses how he has to adapt his posture to perform the exercises accurately. A progress bar gives feedback to the user about how close he is to completing the task.
The video-game like nature of the Kinect itself and the graphical user interface designed with the use of avatars give a playful impression and seek to enhance the patient's motivation to perform his exercises.

To therapists frontend for exercise creation uses the same body-tracking input method, along with a mouse and keyboard. The latter are needed only to name exercises and to determine the number of repetitions required. An exercise can be broken down into multiple components: the initial posture, one or more trajectories and the final posture. In order to compose an exercise the therapist needs to declare a set of postures,. These have to either be recorderd by the tracking device and the therapist performing them himself, or be chosen from preexisting ones. Finally they can be combined into one or more exercises, resulting in a complete session. The allowed error margins for the user can be changed over time, making the same exersices adaptable to the user's growing capabilities.
The setup shows to be non-invasive and motivational while maintaning a high level of comfort to the patient. He can use it independently at home without having to setup and wear further measurement devices. The therapist on the other hand can track the patient's progress remotely and adapt exercises accordingly.

A more complex application of a depth camera in telerehabilitation is shown in [6]. Focussing on the rehabilitation of stroke patient's with impaired hand movement, a refined system is required to correctly track and analyse palm and fingertip positions.
Already existing systems include the Virtual Glove, which tracks up to four IR lights that have to be attached to the user's fingers. This system, like other sensor based tools, is very precise and reliable. But it is also expensive, needs high maintenance, is so delicate that it might break by common use and, in a lot of cases, needs assistance for the stroke patient to even put on.
Visual based systems on the other hand need heavy computation for background and skin extraction, making them less reliable and less efficient in performance.
The proposed tracking technique using the Kinect attempts to avoid the downsides of both of these approaches, while maintaining a high reliability of hand tracking. By definition of the Kinect's depth map, the foreground identification is an easy task that requires no additional computational ressources. It is also not prone to variations of lighting in the scanned evironment. Using a brute force algorithm the location of the palm is deduced by looking for the maximum circular area inside the tracked hand contour. The points with the largest distance to the palm are then interpreted as fingertips, while their connection to the palm represent the fingers themselves.

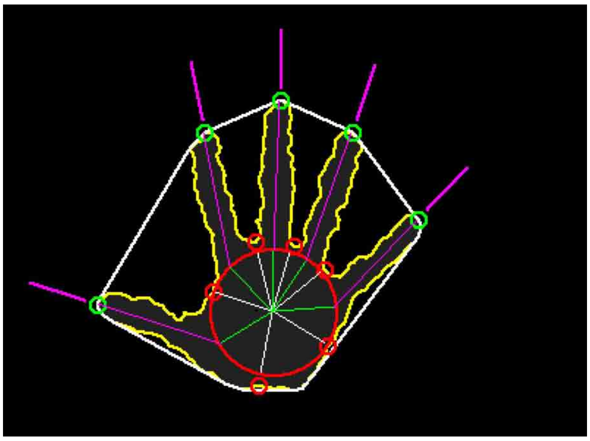


Figure 5. Palm and Fingertip identification
[6]

This technique proves to be more reliable than the Virtual Glove (which can only track up to 4 fingers) only if the hand is facing the sensor. The results are noisy and unreliable in comparison, if the hand is rotated (see Figure 6). This is due to the self occlusion of the different parts of the hand. For a wide range of finger positions and orientations, tested for accuracy with a group of healthy individuals, the Kinect was able to perform well, while failing at tightly closed pinch gestures. The upsides of the Kinect compared to the Virtual Glove may

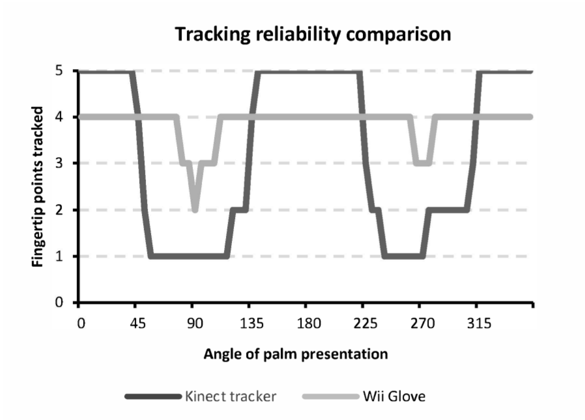


Figure 6. Kinect and Glove comparison
[6]

still outweigh its inaccuracy flaws, but studies with actual stroke patient’s, that may show more unusual gestures due to their disabilities, have yet to be conducted.

Gait Measurement

Human gait is an important indicator of health in diagnosis, monitoring and rehabilitation. Typical systems include setups that require a laboratory setting or expensive and intrusive in-home devices. Depth cameras provide a non-intrusive and markerless way for in-home gait measurement that can yield reliable stride parameters while maintaining the patient’s privacy.

In [7] Stone and Skubic evaluate an in-home gait measurement system using the Kinect and comparing it to two existing methods: a web-camera based system and a marker based motion tracking system.
The marker based motion capture system gives highly accurate results and serves as ground truth data.
The web-camera based system uses two web cameras to monitor the patient, extracting only shilouettes of the body, preserving the patient’s privacy. The shilouettes are computed using a background substraction technique based on color and texture features. The extraction runs at five frames per seconds in real time. The web camera based system is currenly deployed in living facilities for elderly people and has proven to give reliable results for fall detection and gait measurement.
For the depth camera setup two Kinects are used and placed on two opposing ends of the side of a room, facing inwards. One of these sensor serves to evaluate the robustness of the results in respect to camera placement.
The gait parameters are extracted as follows: For the web-camera setting a preexisting algorithm is used that is not further elaborated on, but has already been deployed, as mentioned before. For the Kinect multiple computational steps are nessecary.

Calibration
To obtain useful distance data, a calibration of the Kinect is needed. Using a checkerboard placed on the ground as training setup, the parameters to a real distance formula can be determined.

Foreground extraction
By nature of the obtained depth camera values, the foreground extraction is intuitive and efficient. A simple threshold is used to determine whether a point is located in the background or in the foreground. The background is then easily substracted from the image. The simple and lighting invariant foreground extraction is a significant upside of the depth camera compared to visual devices.

Computation of the Gait Parameters
The centroids of the point clouds given by the depth values of consecutive time frames are projected onto the ground and their spatial differences are computed to determine direction of travel (DOT) and distance travelled. The number of steps is estimated using only data with a certain height from the ground (up to 50cm). A formula for a correlation coefficient is presented, whose local minima indicate right foot steps, and whose local maxima indicate left foot steps. Given the data for the left and right foot steps, the stride time can be deduced, being the time between successive footfalls of the same foot. The stride-to-stride variation can be an important indicator of falls.

Their evaluation of the obtained gait parameters of the Kinect compared to the web-based system and the ground truth data yields mixed results. For one, the placement of the Kinect does have an effect on the accuracy of the results, but the root of the deviations is to be evaluated and remains unclear. The Kinect also showed difficulties detecting clothing that contained a certain amount of spandex. That being said, with a maximum absolute percentage difference of 7.2% from the ground truth values, the Kinect still shows a rather accurate and solid performance. The web-camera approach yielded more accurate results, probably due to its abliity to directly detect foot steps, as opposed to estimating them based on distance traveled. A fusion between RGB camera and depth camera is suggested for further research to further reduce inaccuracies. In [8] Stone and Skubic test their Kinect setup in independent living facilities and expand it to identify walking patterns and distinguish data from different individuals. The results show effective detection of changes in population, but limited indication of actual fall risk due to variation and noise in the measurement. This technique has yet to be refined to be able to detect early signs of changes in health.

In [2] Gabel et al. further expand on Stone and Skubics setup in [7] with the Kinect depth camera as a low-cost, non-intrusive and markerless gait measurement device.
They test a gait parameter extraction method, which is based on the Kinect SDK’s built in skeletal tracking, against ground truth values obtained by an in-shoe pressure sensor and a gyroscope attached to the patient’s wrist.
They suggest a supervised learning process to establish a model that can automatically predict stride parameters.

A few properties have to be extracted from the collected data in order to obtain usable training and testing data. One of these is the direction of progress (DOP). It is derived using the tracked body joints to compute the spatial differences between consecutive centers of mass (COM). Another one is the speed of walking, which equals the norm of the vector that represents the DOP. Calculating the differences of the joint positions themselves and the COM of consecutive frames finally yields a feature vector. This feature vector is then indicative of a certain recorded movement.
Feature vectors have to be computed for all the sensors. A regression tree model is then trained with the feature vectors of the ground truth sensors (i.e. the pressure sensor and the gyroscope). Once the model has been set up correctly, it can predict certain values of interest using the test data obtained by the Kinect sensor.
For stride evaluation these values can be whether the heel or the toe of a person touches the ground, or if the foot is being swinged in the air.
Part of the ground truth data, that has not been used for the learning process, is then compared in accuracy to the test data predictions.
Gabel et al. find out that the resulting measurements are not only accurate, but also robust to the movement of the sensor itself. Their method can be further extended to not only measure foot stance parameters, but also other desired properties like lower limb angular velocities and core posture.

Fall Detection

One of the uses of gait measurement in health care is fall detection. Fall detection is especially important in stationary environments or when caring for the independent living elderly.
Zhang et al. evaluated a fall detection technique using the Kinect depth camera in combination with an RGB camera for their setup. As opposed to Stone and Skubic, Zhang et al. chose to use the built in skeletal tracking model of the Kinect SDK [9].
Their approach is based on width-height ratios of the given joints and their relation to certain body positions. Of the 20 available body joints, only eight were used in order to avoid unnessecary noise. The logarithmic structure difference cost of these eight joints, in any given frame, is visualised in a histogram.
In the resulting graphs they are able to distinguish five different activities in the diagram: standing, fall from standing, fall from chair, sit on chair and sit on floor. The differences between the visualisations of these positions show to be significant enough in order to determine fall positions.

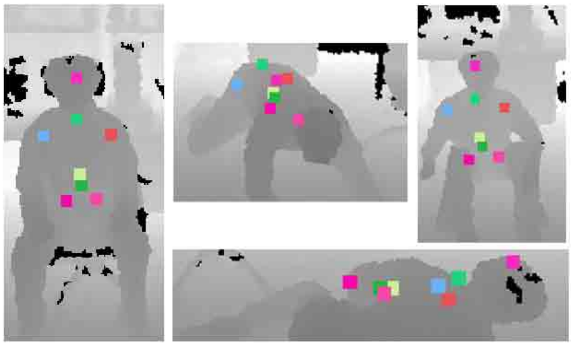


Figure 7. Tracked body joints in different postures
[9]

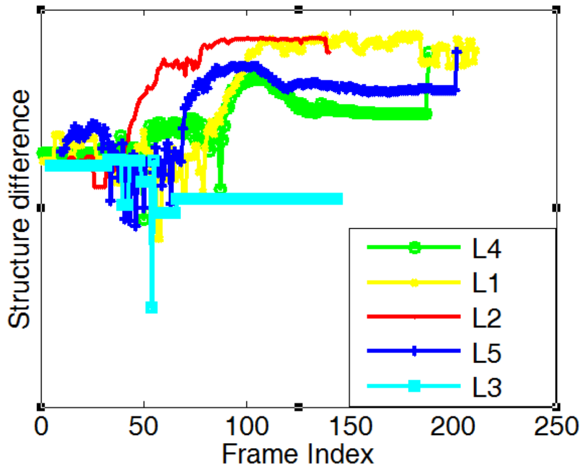


Figure 8. Histogram visualisation of structure differences
[9]

As soon as the person moves out of range of the depth camera, the RGB camera is being deployed. To acquire the values of interest of the recorded data, an additional background substraction has to take place. This is done by computing the changes in consecutive frames, obtaining the area where movement has taken place. The resulting bounding box of the movement area then yields the desired width and height ratios.

Comparing both of these methods, the model built using the Kinect shows to be robust and faster than the model used by the RGB Camera. While the RGB camera is able to detect movement in a wider range, the Kinect performs better when insufficient illumination is present due to its invariance towards changes in lighting.
The success of the setup motivates to expand the model to recognise further activities like taking medicines and human interactions in future work.

DIAGNOSIS

Depth cameras can not only been used as input interfaces in rehabilitation or monitoring devices in-home systems, but also as a tool for diagnosis. An example is Krueger et al.'s research of sleep detection in [3].
Sleep is a significant part of a persons life, either improving their daytime performance or affecting it in a negative way. To diagnose sleeping disorders the patient usually has to stay at special sleep laboratories. For a more comfortable experience, some attempts have been made to allow an in-home evaluation of the patient's sleep. These all include wearable

devices that can measure a persons movement which helps reach a certain diagnosis.
Krueger et al. attempted to further improve the patient's comfort during the evaluation by using the Kinect depth camera to determine movement during sleeping hours.
In their test environment the sensor is placed above the bed, facing down, and records depth values of the persons body during the whole night. From the differences of the depth values over time there are four characteristics that can be derived: the total duration recorded, the number of minutes the patient was identified as awake, the number of minutes the patient was identified as asleep, the time the patient first fell asleep and the percentage of minutes per night that the patient spend sleeping.
Having obtained these characteristics a diagnosis can be reached in accordance to the standardized manual of sleep stages of human subjects [4].
To test the accuracy of their results, Krueger et al. mounted the depth camera over the bed of a patient in a sleeping laboratory, simultaniously collecting their test data and the laboratory data as ground truth.
Their results show an accuracy ranging from 84% in the first test night to 94% in the second. This serves as a prove of concept that sleep diagnosis can in fact be performed unattended at the patient's home, without him having to wear professional equipment during the sleep.

CONCLUSION

It has been shown that depth cameras can be widely applied in health care to operate software, to monitor patient's and even to diagnose disorders. While a lot of the exemplary applications have existing solutions, depth cameras provide undeniable benefits to all these processes.

During surgical procedures, where it is important to keep the work environment clean and intuitive, a touchless interface is the most fitting solution. Gestural input is easy to use and does not require cleansing of additional contact devices. Furthermore the needed space is minimised. Depth cameras as touchless interfaces are comfortable to use, as opposed to wearable devices, not requiring any physical contact. The accuracy of the Kinect in this setting has shown to be acceptable, but a device like the Leap Motion Sensor, that is purchasable at about the same low cost as the Kinect, might be preferrable. The wider interaction range of the Kinect can be viewed as a positive, allowing the user to operate the device from further away, but it can also lead to involuntary inputs. The Leap Motion Sensor in contrast yields a higher accuracy and has a much smaller interaction area.
While this short sensoring distance is of use in the case of operating surgical software, a wider distance is needed for settings such as in-home rehabilitation. In-home rehablitation, as mentioned before, is a huge benefit to health care, reducing costs and face-to-face treatment time, as well as increasing the patient's comfort. Telerehabilitation software based on touchless interfaces like KiRes succeed in providing a motivational, game like experience for the patient while still monitoring his progress. Although still expandable, KiReS shows no downsides in in-home rehabilitation. Especially for patient's with disabilities, as is the case with stroke patient's, depth cameras are preferrable to wearable sensor devices, as these patient's may struggle putting devices on and as they may experience discomfort while using them. Here it is important to try and maintain the patient's quality of life and motivation. This does not apply to exersices that require a detailed movement comprehension of small body parts like the hand. In this case the Kinects conveniance cannot make up for its measurement inaccuracies in comparison to the Virtual Glove. Still, the Virtual Glove is a very expensive, delicate, and hard to put on devise that is rather not being used without supervision.
As a monitoring device, the depth camera offers improvements in foreground extraction, lighting invariance and robustness. While other methods using visual cameras as well as wearable sensors yield more accurate results, the differences of the data obtained with the Kinect are usually low and still reliable.

All in all, compared to existing setups, the Kinect as examplatory depth camera is a low-cost solution that offers robust and for most purposes reliable body tracking. Its maintanance is low since no batteries need to be exchanged and the setup is simple and needs no additional supervision. Therefore it can be used as an in-home tracking system, as opposed to most other devices that require a laboratory setting. The depth tracking technique is invariant to changes in lighting which makes it usable even at night. By obtaining depth data only it speaks to the privacy concerns of the monitored patient's. By originally being part of an entertainment system patient's are less reluctant to use the device at home. Although the mentioned applications still need to be refined, they are proof of concept that depth cameras can be a viable low-cost opportunity for in-home monitoring, rehabilitation and diagnosis in health care.

FUTURE WORK

While proof of concept in all the presented applications has been given, wider ranged evaluations are a nessacity to make a definitve statement about their usability.
In some cases the proposed setups might be usable as shown with small adjustments based on further evaluations. In other cases, for example the tracking of hand impaired patient's, the technology of the used depth camera just doesn't deliver the needed information. A similar setup with the Leap Motion Sensor instead of the Kinect 1.0 might yield better results, while still being low-cost and convenient.
But the Leap Motion Sensor is not the only good alternative anymore; Since the publication of the presented projects the Kinect 2.0 has been released. It is based on TOF measurements, which have a wider range and a higher accuracy than IR based sensors. TOF cameras are usually very expensive, but like its predecessor the Kinect 2.0 comes at a low and affordable price, making it viable for wider applications.
Using a TOF based camera, range restrictions can be overcome, making monitoring in bigger rooms and even outside areas possible.
As of in-home interaction devices, it has been rumored that Apple is going to release an iPad with an integrated 3D sensoring device. The integration of a 3D scanner into an item

that already sees everyday use by many people, is surely going to increase the patient's acceptance, when used for health care. The underlying technology is yet to be disclosed, but might just open up new and better ways of motion tracking.

REFERENCES

1. Antón, D., Goñi, A., Illarramendi, A., Torres-Unda, J. J., and Seco, J. Kires: A kinect-based telerehabilitation system.
2. Gabel, M., Gilad-Bachrach, R., Renshaw, E., and Schuster, A. Full body gait analysis with kinect. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE* (Aug 2012), 1964–1967.
3. Krüger, B., Vögele, A., Herwartz, L., Terkatz, T., Weber, A., Garcia, C., Fietze, I., and Penzel, T. Sleep detection using a depth camera. In *Computational Science and Its Applications–ICCSA 2014*. Springer, 2014, 824–835.
4. Rechtschaffen, A., and Kales, A. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects.
5. Ruppert, G., Reis, L., Amorim, P., de Moraes, T., and da Silva, J. Touchless gesture user interface for interactive image visualization in urological surgery. *World Journal of Urology 30*, 5 (2012), 687–691.
6. Shires, L., Battersby, S., Lewis, J., Brown, D., Sherkat, N., and Standen, P. Enhancing the tracking capabilities of the microsoft kinect for stroke rehabilitation. In *Serious Games and Applications for Health (SeGAH), 2013 IEEE 2nd International Conference on* (May 2013), 1–8.
7. Stone, E., and Skubic, M. Evaluation of an inexpensive depth camera for in-home gait assessment. *Journal of Ambient Intelligence and Smart Environments 3*, 4 (2011), 349–361.
8. Stone, E. E., and Skubic, M. Unobtrusive, continuous, in-home gait measurement using the microsoft kinect. *Biomedical Engineering, IEEE Transactions on 60*, 10 (2013), 2925–2932.
9. Zhang, C., Tian, Y., and Capezuti, E. *Privacy preserving automatic fall detection for elderly using RGBD cameras*. Springer, 2012.
10. Zhang, Z. Microsoft kinect sensor and its effect. *MultiMedia, IEEE 19*, 2 (Feb 2012), 4–10.

Non-optical remote object recognition and tracking in Human-Computer Interaction

Michael Götze
Mähringer Weg 130
89075 Ulm
michael.goetze@uni-ulm.de

ABSTRACT

In this paper recent non-optical techniques for remote object recognition and tracking are introduced. These techniques presenting alternatives to common optical recognition and tracking methods, they're less or completely not impaired by occlusion or illumination changes. Furthermore their benefit to HCI research topics like accuracy, distance, performance and the limitation of interaction space as well as their restriction are presented.

INTRODUCTION

Recognition and tracking in Human Computer Interaction (HCI) is mostly based on optical techniques such as infrared or depth cameras. Therefore users need to be in the camera's sight of view to interact with the system. Illumination changes or occlusion makes it impossible for such systems to track objects and recognize performed gesture. In addition, mostly mobile, devices face the problem of a small limited interaction space, like displays of smartphones or -watches. In this paper recognition and tracking methods are presented which are avoiding exactly these problems instead of using classic optical methods.
These implementations are presented in seperated chapters relating to their utilized sensors. Furthermore the presented techniques are evaluated relative to their gain to HCI topics such as distance and accuracy.

RECOGNITION AND TRACKING TECHNOLOGIES

Magnetic sensing

A usually embedded sensor in current mobile devices is the magnetometer (compass), which can be used to increase interaction space of mobile devices, e.g. providing Around Device Interaction (ADI) for the recognition of 3D movement gestures [12].

MagiTact uses the magnetometer for ADI in combination with a magnetic material which can be hold in the hand (e.g. pen, ring, etc.). Performing gestures with these materials in 3D space around the device influence the values of the magnetometer along x, y and z axis. But these values are also affected by different magnetic fields around the device. The earth's magnetic field is the biggest influence. To get more accurate values, the influence of these factors has to be decreased. Hamed Ketabdar et al. achieved this by calculating a derivative of the received values over time using an individual designed high pass filter. This high pass filter subtracts two successive values to get solely relevant value changes, which belongs to the performed gesture.
The high pass filtered values has to be further processed with feature extraction by the following features, which are "average strength of magnetic field in different directions, average piecewise correlation between field strength in different directions, and zero crossing rate (for different directions)" [12]. To identify begin and end of any gesture the change of signal dimension over a threshold is considered. All features are extracted from begin to end of the gesture. For the final gesture grading the extracted features are classified in a heuristically designed binary decision tree.
MagiTact recognizes up to six gestures with an accuracy over 90%. It can be used in daily for turning pages, accepting/rejecting calls or controlling a music player while the mobile device is in the pocket [12].

MagiTact is e.g. useful used in *MagiMusic* to imitate the feeling of playing musical instruments such as Air Guitar, Harmonics or Drum Kit [11].

Figure 1. Selection of an menu item with *Nenya*.

Another technique using magnetic sensors is *Nenya*, a finger ring in combination with a wrist-worn sensor (baselet), based

on a 3-axis magnetometer. *Nenya* is a strong permanent magnet. The magnetic poles are located on opposite sides of the ring. Additionally a small disc magnet is added to explicit show the rings position to the user. The baselet contains a 3-axis magnetometer to track the position of *Nenya*. Via Bluetooth the baselet can transmit *Nenya*'s input to another device. By twisting the ring as shown in Figure 1, a 1D parameter is entered. A selection can be confirmed by sliding *Nenya* along the finger. When *Nenya* is spinned, the magnetic field changes. These changes are sensed by the magnetometer in the baselet.
The software developped by Daniel Ashbrook et al considers only the x- and z-axis to determine the angle relative to finger at which *Nenya* is being spinned. In order to detect a selection, the x- and y-axis are regarded. By moving *Nenya* away from the baselet the measured magnetic field strength decrease, if it drops under a threshold a selection is detected by the baselet.
Nenya can be used as a discret input device in combination with a wrist-worn magnetometer which is e.g. included in smartwatches. For sure it's limited by the amount of selectable targets [1].

Another approach to use electric field sensing is the implementation of *Geremin*. It uses one antenna behind a steering wheel in a car for 2D gesture recognition. *Geremin* uses a modified Theremin, an electronic music intrument, invented by Professor Leon Theremin in 1928, which can be controlled without any physical contact and is composed of two metal antennas. If the hand is moved towards or away from the installed antenna the capacity of an oscillating circuit changes. They used the tool Praat to fed this generated sound into a signal processing component that translates the pitch curve into a vector of numbers. This feature vector is used as input for gesture recognition. Christoph Endres et al. used the 3D gesture recognition system for multimodal dialog systems [18] developped by Neelrath, Robert, and Jan Alexandersson in combination with Multi-Dimensional Dynamic Time Warp (DTW) [23] algorithm, for gesture classification. *Geremin* recognizes 10 gestures and six of them (left and right, up and down, anticlockwise circle and square) with an accuracy more than 70% [4].

Audio sensing

Like magnetic field sensing, the requirements for audio sensing is available at nearly every mobile device. The below presented techniques are using build in speakers and microphones for gesture detection [9].

SoundWave uses the doppler effect for gesture recognition. The doppler effect is a frequency shift of a sound wave, which is proportional to the moving objects speed and source frequency. The embedded speakers of an laptop or another mobile device create an inaudible continuous tone between 18-22 kHz. Every motion up to a range of one meter, will effect doppler-shifted reflections. The included microphones record the reflected signal of a moving object, to valuate motion and gestures through the monitored frequency variations.
Sidhant Gupta et al. buffer the incoming signal from the microphone and estimate the Fast Fourier Transformation (FFT). The estimated FFT vectors are further processed by their evolved filters, signal conditioning, bandwidth extraction, motion detection and feature extraction, to determine if a gesture was performed. *SoundWave* detects four hand motions (seesaw, toward or away, pull-back, double tab, slow tab) with an accuracy of more then 90%. It's even possible to play music on the same device without harming the performance of *SoundWave* [9].

SoundWave's approach is reused in *AirLink*, which allows the user a multiple-device environment (MDE) e.g. to share files between them. By waving the hand from the initialising to the receiving device. *AirLink* detects three gestures and represent each as a unique codeword, hand moving towards (T), away (A) as well as towards and then away (X) from the device. With these motions *AirLink* can identify the relative position of the devices and select the interacting ones like shown in Figure 2 [2].

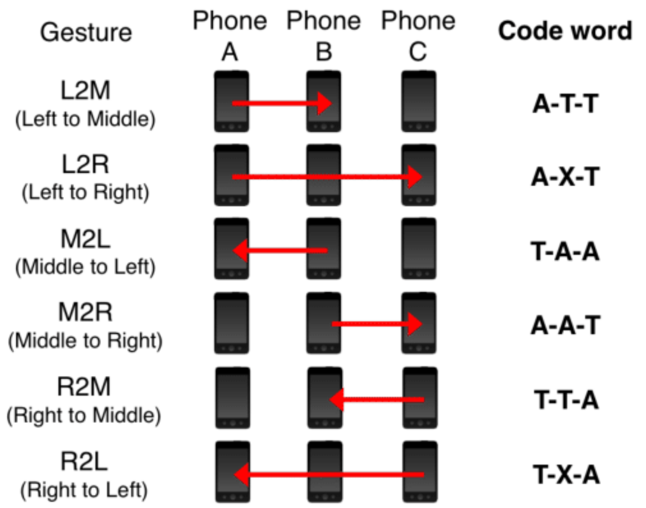


Figure 2. The combination of 3 basic hand movements: T (toward), A (away), and X (toward and then away), represents each type of gestures as a unique codeword.

Another implementation of a MDE is *SurfaceLink*. It enables the detection of devices on the same surface as well as the recognition of gestures performed on that surface.
SurfaceLink uses accelerometers, vibration motors, speakers, and microphones, which are yet integrated in most mobile devices. Mayank Goel et al. use the fact that hard, flat surfaces conduct vibration very well, to detect if the devices are placed on the same surface, by either user-induced (knocking) or device-induced (virbration motor) vibrations.
The user-induced vibrations can be sensed by the on-device accelerometers. At first, the host-device needs a list of nearby devices to instruct them to start sampling their accelerometers at 100Hz. The required list is generated by GPS and Wi-Fi information.
The devices share their measured accelerometer data and pass it through a high-pass filter. Between these data samples, a pairwise cross-correlation is performed, to cluster the devices in two groups, either on the same surface or not.
The device-induced vibrations are coupled by the air and can be heard by an on device microphone. To pick up these subtle

vibrations, noice cancellation, using two microphones (one touching the surface and one in the opposite direction), is required. In the same way, like in the user-induced case, the device retrieves a list of nearby devices and instruct them to start sampling their microphone at 44.1 kHz.
When a device vibrates on a hard, flat surface, it generates a low frequency sound. The system subtracts the audio from the top from the microphone touching the surface for noise cancellation. The devices share their data and a pairwise cross-correlation and clustering is done like in the user-induced case.
SurfaceLink detects gestures using the fact, that dragging a finger/hand over a hard, flat surface is generating vibrations, which can be measured by the on-device microphone. The generated sound of every gesture has different characteristics in different directions (e.g. louder when near the device). Each device classifies the gesture into four gesture properties depending on the measured sound: Gesture Class (away, towards, non-participating, pinch, expand, fast swipe), Gesture Length (quarter, half, full), Touch Mode (fingertip, fingernail, fist) and Gesture Shape (line, polygon, triangle, circle, semicircle). Depending on distance between the devices and the performed gesture, the observed energy differs. Each device makes a decision once for each gesture and sends the classification result, classification confidence, and the total observed energy to the host device. To decide which devices were supposed, the host-device evaluates the classifications from the other devices. In case there is more then one device predicting to be meant, the host-device checks the confidence classification and if needed the observed energy for the highest results [5].

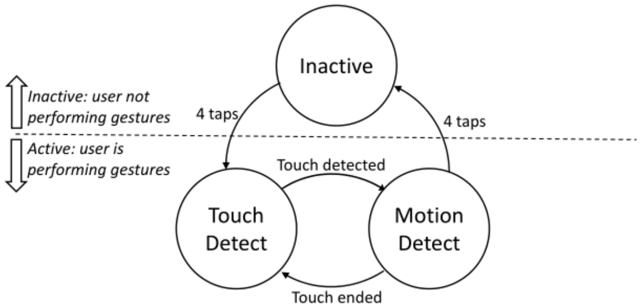


Figure 3. The three states of the Energy Harvesting Wearable Ring.

Jeremy Gummeson et. al developped another approach according to the idea behind *SurfaceLink*, the *Energy Harvesting Wearable Ring*.
The ring includes an accelerometer, a tendon force sensitive resistor (FSR) and a microphone. These are the three main sensing components. The accelerometer is used for gesture identification by recording inertial data during a performed gesture. The FSR detects if the finger touches a surface and the microphone is spent to detect the motion of the finger, based on the generated audio by the fingers motion on the surface [6].
For energy harvesting they used a NFC chip which passively recharges the battery while the ring is hold next to a phone [7]. Additionally while the ring is inactive, it's accelerometer enters a low-power mode. The implemented states are shown in Figure 3. To activate the ring for input, the user has to tap a surface four times, the ring enters Touch Detect state and activates the FSR. By touching a surface the ring enters Motion Detect state and the microphone is turned on. At the end of motion or the touch, the measured data from the microphone and the accelerometer is fed to the classifier on the ring, to identify the gesture. As soon as a gesture is identified, the input is wireless transmitted to the remote device. By tapping four times on a surface, while the ring is active, it turns inactive [6].

Inertial sensing

The combination of multiple inertial sensors improves the amount of recorded data and can lead to a higher accuracy in gesture recognition. *Pingu* is a further *smart* finger ring, which is equipped with an accelerometer, magnetometer, gyroscope, proximity sensor and Bluetooth adapter. *Pingu* can recognize performed gestures in the air, on top of a table or on the palm. The 3-axis accelerometer detects the orientation and fluctuation of the ring along the x, y and z axis. A 3-axis gyroscope is used to detect the angular rate of the ring's movement along the three axis.
The combination of accelerometer and gyroscope provides six degree of freedom, what is used to detect the 3D trajectories of *Pingu*. Moreover the magnetic field sensor measures the deformation of magnetic fields to recognize coarse gestures which are performed around the device. Additionally the proximity sensor is installed to sense the proximity of the other fingers and the Bluetooth adapter is used to transfer the input to another device [17].
Mehran Roshandel et al. fed the collected data of all sensors to their classification algorithm which is based on the implementation of the Weka machine learning toolkit [24]. *Pingu* got a set of nine gestures including cross, swipe (right and left) and circle (in opposite directions) gestures [17].

In the approach, Christopher-Eyk Hrabia et al. used 8 Magnetic Angular Rate Gravity (MARG) sensors placed on a hand. Each MARG sensor consist of accelerometer, magnetometer and gyroscope for joint orientation tracking. They've shown that the amount of 8 MARG sensors, placed at the right location, is sufficient to detect hand motions, finger joint flexion angles and motion paths. The collected data of the sensors is logged by a microcontroller and further processed by their developped Java-based PC-Application. Their application uses the Mahony's [16] complementary filter in combination with Madwig's [15] magnetic distortion filter to fuse the sensor data and compute the 3D orientation. A whole hand model is generated by the software and the logged sensor data is used for motion detection [10].

John Sunwoo et al. developped a more minimalistic approach including only one 3-axis accelerometer placed at a gesture recognition band which is worn at the wrist. Additionally a IrDA transceiver, for interaction with other devices via bluetooth and WLAN, is installed. The approach implies their implemented software recognition engine that receives and recognize the gesture commands. Each gesture is classified by the gesture recognition engine according to the segmentation shown in Figure 4. Overall it recognizes a set of 12 gestures with an accuracy of 97,6% by a well adapted user [3].

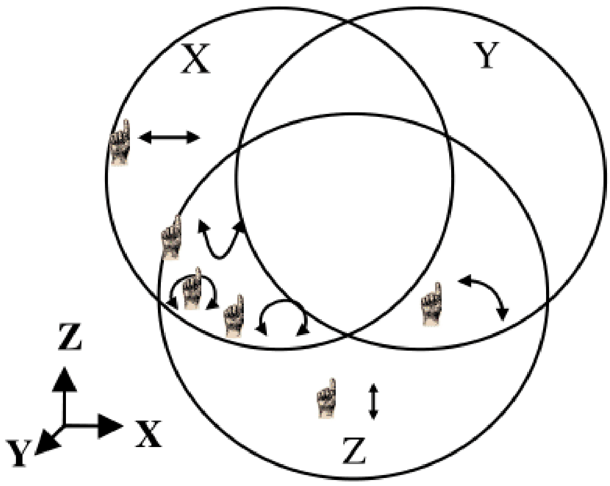


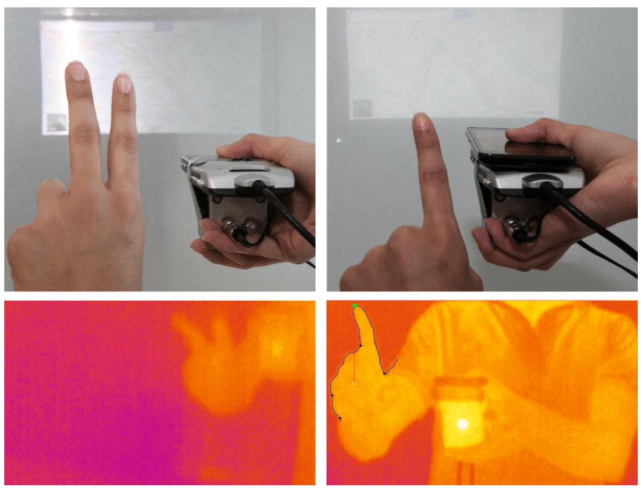
Figure 4. Segmentation diagram of gesture commands.

Other sensing technologies

LightWave is a system that doesn't profit from sensors which are usually embedded in mobile devices, but it present another possible usage of a sensor which is yet available in a lot of homes and workplaces.
Sidhant Gupta implemented *LightWave*, which turns unmodified fluorescent light (CFL) bulbs into sensors of human proximity and allows hover gesture (waving a hand close to a lamp) recognition near a CFL.
The proximity of a human body to a CFL bulb causes predictable variations in electromagnetic noise resulting from the change in impedance. These variations in the electromagnetic interference can be measured over time, by sampling the power line at appropriate frequency. This frequency can be detected, from *LightWave*, by turning the lamp on and off. Therefore only a single interface device plugged into any electrical outlet is required [8].

Another approach extending the usability of an, in homes and workplaces, available object, is *WiSee*. It uses the doppler effect of wireless signals for gesture recognition to transform OFDM-based systems (802.11 a/g/n, WiMAX, LTE, etc.) into a sensor. A mobile device like a phone or laptop can be used as signal source, hence *WiSee* is not effected from occlusion and even works e.g. in another room. The human body acts as virtual transmitter and reflects the signals.
If e.g. an object is moved forward the signals arrives faster. Performing a gesture, everywhere in the range of the wireless router, results in a pattern of doppler shifts at the wireless receiver. These doppler shifts can be detected from the receiver by transforming the received signal into a narrowband pulse with the bandwidth of a few Hertz and track the frequency of this narrowband pulse. *WiSee* detects a set of nine gestures including punching, pushing, pulling and circle gestures.
Qifan Pu et al. used their prototype in application scenarios such like skipping channels on TV, turning volume up and down or controlling the room temperature [20].

Sahami Shirazi et al. showed how to enlarge interaction space with a thermal camera. Thermal cameras are producing thermograms of a surface based on it's incident radiation. This radiation consists of the surface's emitted energy, depending on it's temperature, as well as the radiation reflection of surrounding objects. A surface can act, for thermal cameras, as a mirror if it reflects radiation in a specular manner. Thus, the camera's field of view is enlarged, by the possibility to view objects besides or even behind it, as shown in Figure 5.
To enable this, it's essential to find a proper surface. Surface made of medium-density fiberboards (MDF), different metals or with a smooth paint can be used. Even some materials which are transperant for visual light, such as transparent glass or plastic, can be used for thermal reflection. Sahami Shirazi et al. used Otsus thresholding method to detect relevant parts of a human's body (hand, finger), in the taken pictures [19].
They implemented two methods of gesture recognition, one for interacting with/on the surface and the other for (hover) gestures made in the air. To interact with/on a surface they use heat trace detection. By touching the surface, the user transmits heat or cold. The difference in temperature is detected and the area, shape, and temperature of the contour is examined. The gesture mapping is based on matching the shape of the contour detected [21].
For in-air gesture recognition they utilized the approach of Björn Stenger et al. [22] by tracking fingertips and computing their relative distance. This approach can be used in realtime without any latency or delay. However, the detection of an object needs discrepancy in the temperature of object and focused surface. Their prototype consists of a thermal camera, a pico-projector, and a smartphone. In their usecase the projector streams a map on a surface. The user can zoom in and out [21].



(a) Aluminum (b) Glass

Figure 5. Thermal camera focuse on (a) Aluminium and (b) Glass to recognize the hand besides it.

RECOGNITION AND TRACKING IN HCI

The presented implementations will be differed into different HCI research topics and evaluated relating to their limitations and opportunities.

Tracking-Distance

Depending on far distances, only *WiSee* enabled gesture recognition in a big area (in the range of the wireless router) [20]. Other implementations like *Pingu* [17] or the *Wearable Gesture Recognition Band* [3] enable at least to interact with another device in the range of Bluetooth.
SurfaceLink detects other devices on the same surface with an accuracy of 97,7% when devices are up to 2,43 meters apart.

Accuracy

The approach of *MagiTact* [12] reached an accuracy, of recognizing the correct gesture, over 90%. In addition the user only needs magnetic object such as a ring or a pen to enable tracking to the magnetic field sensor. In opposite to *MagiTact* [12], *SoundWave* [9] uses audio sensing but also utilizes only build-in sensors, the microphone and speakers. *SoundWave* gains an gesture recognition accuracy average over 90%.
At this point *SurfaceLink* [5] should be mentioned. Mayank Goel et al. recognized the precense of a device on the same surface with 99,7%, the set of performed gestures with 90,3% and the total arrangement of the devices with 89,4%.

When to use Non-optical Recognition and Tracking Technologies

Current optical tracking methods are using time of flight or structured light sensing to recognize body parts as well as their motion. These techniques are facing problems to track relaible data in cases of occlusion, illumination changes or reflective surface structures. In cases where it isn't possible to get reliable results or even no results, systems would provit to collect data from non-optical techniques. Additionally, non-optical sensors can be combined with optical ones to improve their accuracy of gesture detection, what's e.g. used in *Wii* [13][14].

DISCUSSION

Recently there are insufficient implementation which enable gesture recognition in a big area. *WiSee* enabled it without facing occlusion problems, but still the demand of such technologies is high, the set of gestures too small and new techniques are desirable for future researchers. *WiSee* faced the near-far problem, in which a receiver captures a strong signal and thereby makes it impossible to detect a weaker signal. Thus, it couldn't detect a performed gesture if a second person was closer to the wireless router. They solved this problem by using more antennas (up to 5), to enable multiple user environment [20].

Currently we got a lot of smart devices but there aren't enough good possibilities to let multiple device interact reasonable interact with each other. *SoundWave* [9] and *SurfaceLink* [5] are two approaches facing this topic and enable MDE.
SoundWave enables in-air gestures with a great performance and accuracy even if the device is placed in an noisy environment where e.g. music is played. The key drawback of *SoundWave* is the generated sound between 18-22 kHz which may can be heard by children or animals. Additionally some devices prevents sound generation and recording over 18 kHz what excludes themselves to utilize *SoundWave* [9].
SurfaceLink detects gestures performed on a surface and enables around device interactions. The rich set of gestures can be very useful for single or multiple device environments. However the utilization of *SurfaceLink* can have significant power implications, cause by the extended use of microphone and accelerometer [5].

For unremarkable user input *Nenya* can be very handy with up to eight selectable targets. In the implemenation of Daniel Ashbrook et al. a bracelet is required. This limitation can be disabled e.g. in combination with a smartwatch, which also includes the required sensors for *Nenya*. Another limitation is obviously that the user has to know which meaning belongs to each of the eight invisible targets [1].

The energy harvesting ring developped by Jeremy Gummeson et al. used a NFC chip to passively recharge the battery while the ring is hold next to a phone. The integrated battery enables more than 10 hours of active user input. However, they need to improve their gesture recognition performance. Maybe the factor of having no direct user feedback impaired the correct gesture learning process of the users. The evaluation of produced audio data from other surfaces wil also improve the gesture classification performance [6].

Using thermal cameras to enlarge the tracking and interaction space shows a promise approach for future researchers but for the moment smaller and cheaper thermal cameras are needed to integrate them in common mobile devices. Sahami Shirazi et al.'s approach allows to use either surfaces which diffuse visual light or are transparent for visual light as a mirror for thermal reflection. If the object and the surface/background have a similar temperature, there is no difference between the object the background in the thermal image. In this case it's very hard to detect the object and it's performed gesture. Furthermore not every surface is appropriate, the surface should be smooth and polished to acquire a good thermal reflectivity [21].

CONCLUSION

The preseneted technologies have shown that there are alternative ways to track objects and recognize gestures without using optical methods such as depth cameras and infrared. These methods aren't affected by occlusion or illumination changes and most of them are using build-in sensor or transform available tools like microphones to sensors. We can profit a lot without using new sensors with higher costs, we just can enlarge the utilization of available sensors.
Interacting at home without wearing any tool will facilitate everyone's everyday life. Therefore the usage of wireless router as whole home gesture recognition tool, based on *WiSee*'s approach, will definitely be important for HCI researchers in the future.
A lot of developer are looking for small, unremarkable devices just like the rings [1][6][12][17], presented in this pa-

per. Recently they aren't small and accure enough but in the future, integrated sensors will be smaller and cheaper to be interesting as a interacting device in everyday life.
Gesture recognition with inertial sensors, e.g. integrated in the *Wearable Gesture Recognition Band* [23], could be easily implemented in smartwatches, which yet have integrated inertial sensors. The combination of inertial sensors and mircrophones shown in the approach of Jeremy Gummeson et al. [6] or *SurfaceLink* [5] is promising and will be hopefully component of future researches.
Thermal reflection is also promising in certain environment and usecases. Therefore thermal cameras need to be smaller and cheaper to integrate them in common mobile devices. Additionally the cameras and algorithms need to be optimized to increase thermal reflection and extend the range of surfaces, more gestures and higher accuracy [21].
In my opinion, the combination of optical with non-optical tracking techniques can be very usefull to increase their accuracy and reliability. This could be e.g. reasonable implemented by adding a gesture recognition system like *WiSee* to kinect. Thereby occlusion problems could be eliminated.

REFERENCES

1. Ashbrook, D., Baudisch, P., and White, S. Nenya: Subtle and Eyes-Free Mobile Input with a Magnetically-Tracked Finger Ring. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*, ACM Press (New York, New York, USA, May 2011), 2043–2046.
2. Chen, K.-Y., Ashbrook, D., Goel, M., Lee, S.-H., and Patel, S. AirLink: Sharing Files Between Multiple Devices Using In-Air Gestures. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '14 Adjunct*, ACM Press (New York, New York, USA, Sept. 2014), 565–569.
3. Cho, I.-Y., Sunwoo, J., Son, Y.-K., Oh, M.-H., and Lee, C.-H. Development of a single 3-axis accelerometer sensor based wearable gesture recognition band. In *Proceedings of the 4th International Conference on Ubiquitous Intelligence and Computing*, Springer-Verlag Berlin Heidelberg (July 2007), 43–52.
4. Endres, C., Schwartz, T., and Müller, C. A. Geremin: 2D Microgestures for Drivers Based on Electric Field Sensing. In *Proceedings of the 15th international conference on Intelligent user interfaces - IUI '11*, ACM Press (New York, New York, USA, Feb. 2011), 327.
5. Goel, M., Lee, B., Islam Aumi, M. T., Patel, S., Borriello, G., Hibino, S., and Begole, B. SurfaceLink: Using Inertial and Acoustic Sensing to Enable Multi-Device Interaction on a Surface. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, ACM Press (New York, New York, USA, Apr. 2014), 1387–1396.
6. Gummeson, J., Priyantha, B., and Liu, J. An energy harvesting wearable ring platform for gestureinput on surfaces. In *Proceedings of the 12th annual international conference on Mobile systems, applications, and services - MobiSys '14*, ACM Press (New York, New York, USA, June 2014), 162–175.
7. Gummeson, J. J., Priyantha, B., Ganesan, D., Thrasher, D., and Zhang, P. Engarde: Protecting the mobile phone from malicious nfc interactions. In *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '13, ACM (New York, NY, USA, 2013), 445–458.
8. Gupta, S., Chen, K.-Y., Reynolds, M. S., and Patel, S. N. LightWave: Using Compact Fluorescent Lights as Sensors. In *Proceedings of the 13th international conference on Ubiquitous computing - UbiComp '11*, ACM Press (New York, New York, USA, Sept. 2011), 65.
9. Gupta, S., Morris, D., Patel, S., and Tan, D. SoundWave. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, ACM Press (New York, New York, USA, May 2012), 1911.
10. Hrabia, C.-E., Wolf, K., and Wilhelm, M. Whole hand modeling using 8 wearable sensors. In *Proceedings of the 4th Augmented Human International Conference on - AH '13*, ACM Press (New York, New York, USA, Mar. 2013), 21–28.
11. Ketabdar, H., Jahanbekam, A., Yuksel, K. A., Hirsch, T., and Haji Abolhassani, A. MagiMusic: Using Embedded Compass (Magnetic) Sensor for Touch-less Gesture Based Interaction with Digital Music Instruments in Mobile Devices. In *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction - TEI '11*, ACM Press (New York, New York, USA, Jan. 2011), 241.
12. Ketabdar, H., Yüksel, K. A., and Roshandel, M. MagiTact: Interaction with Mobile Devices Based on Compass (Magnetic) Sensor. In *Proceedings of the 15th international conference on Intelligent user interfaces - IUI '10*, ACM Press (New York, New York, USA, Feb. 2010), 413.
13. Langmann, B., Hartmann, K., and Loffeld, O. Depth camera technology comparison and performance evaluation. In *ICPRAM (2)* (2012), 438–444.
14. Lin, J., Nishino, H., Kagawa, T., and Utsumiya, K. Free hand interface for controlling applications based on wii remote ir sensor. In *Proceedings of the 9th ACM SIGGRAPH Conference on Virtual-Reality Continuum and Its Applications in Industry*, VRCAI '10, ACM (New York, NY, USA, 2010), 139–142.
15. Madgwick, S. O., Harrison, A. J., and Vaidyanathan, R. Estimation of imu and marg orientation using a gradient descent algorithm. In *Rehabilitation Robotics (ICORR), 2011 IEEE International Conference on*, IEEE (2011), 1–7.
16. Mahony, R., H. T. P. J. Nonlinear Complementary Filters on the Special Orthogonal Group. In *Automatic Control,*

IEEE Transactions on (Volume:53 , Issue: 5), IEEE (2008), 1203 – 1218.

17. Mehran Roshandel, Aarti Munjal, Peyman Moghadam, Shahin Tajik, H. K. Multi-sensor ased Gestures Recognition with a Smart Finger Ring. In *16th International Conference, HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part II*, Springer International Publishing (2014), 316–324.

18. Neßelrath, R., and Alexandersson, J. A 3d gesture recognition system for multimodal dialog systems. *6th IJCAI Worksh. on Knowledge and Reasoning in Practical Dialogue Systems* (2009), 46–51.

19. Otsu, N. A threshold selection method from gray-level histograms. *Automatica 11*, 285-296 (1975), 23–27.

20. Pu, Q., Gupta, S., Gollakota, S., and Patel, S. Whole-home gesture recognition using wireless signals. In *Proceedings of the 19th annual international conference on Mobile computing & networking - MobiCom '13*, ACM Press (New York, New York, USA, Sept. 2013), 27.

21. Sahami Shirazi, A., Abdelrahman, Y., Henze, N., Schneegass, S., Khalilbeigi, M., and Schmidt, A. Exploiting thermal reflection for interactive systems. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, ACM Press (New York, New York, USA, Apr. 2014), 3483–3492.

22. Stenger, B., Thayananthan, A., Torr, P. H., and Cipolla, R. Model-based hand tracking using a hierarchical bayesian filter. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 28*, 9 (2006), 1372–1384.

23. Ten Holt, G., Reinders, M., and Hendriks, E. Multi-dimensional dynamic time warping for gesture recognition. In *Thirteenth annual conference of the Advanced School for Computing and Imaging*, vol. 300 (2007).

24. Witten, I. H., Frank, E., and Hall, M. A. *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2011.

Are we getting unsocial? Exploring Computer Mediated Communication.

Michael Olah
michael.olah@uni-ulm.de

ABSTRACT
Communication is a basic element for living in a society and having a social life. Technology like computers and mobile phones changed the way how people communicate. In earlier days social interaction depends on a Face-to-Face conversation that means to be simultaneously at the same place. Today E-Mail, Instant Messaging, Videoconferences are a few examples how Computer Mediated Communication (CMC) offers more methods to stay and get in contact with other humans over distance. However there is an opinion that this technology makes the people unsocial.
This paper argues that acting unsocial is a problem of human behaviour and not of CMC. It gives an overview of current types of Computer Mediated Communcation and their social effects. Interpersonal interaction needs touching the other as well. CMC has still limits in transfering such tangible communication between two humans. Approaches to overcome these limits are presented.

INTRODUCTION
In daily life you can recognize more and more people sitting next to each other and interact with their mobile device instead of their surroundings. You can find couples who sit in a restaurant and both of them looking on their smartphones and not in the face of the other persons. This situation could be described as „phubbing". The word is compound of „phone"and „snubbing"and was created by the Macquarie Dictionary for the campaign to StopPhubbing [1]. In this campaign they state that it is is unpolite to interact with your mobile device instead of paying attention to the person in the current social setting.
A lot more examples of situations where people prefer interacting with their mobile device instead of their human environment could be listed. This fact could create the opinion, that people are getting unsocial caused by technology.
But is it true that people behave unsocial using CMC or do we have to adapt the meaning of being social? Text messages and E-Mail are popular acitivities on smartphones. US people use the possibility to communicate with known people who are temporarily not next to them [2]. From this point

7th Seminar on Research Trends in Media Informatics (RTMI '15). February 2015, Ulm University, Ulm, Germany.

Figure 1. Common view of low social interaction in public places

of view mobile phones are used for a social interaction. The technology makes it possible to maintain social contacts and to communicate with friends via instant messaging, be part in a online community or social networks. Computer Mediated Communication offers a lot advantages for communication with other people. So why is there the opinion that CMC makes the people unsocial while it offers more and more ways of communication and so being social?
Before talking about communication, computer mediated communication and social issues, a definition has to be given. A common dictionary entry defines communication as an exchange of thoughts, messages or information by speech, signals, writing or behaviour [3]. Another approach is the origin of the latin word „communicare "and its meaning to unite, to share or to inform something or someone. So with a communication you unite two or more subjects and exchange information. But Communication depends not only on the information itself. It is a combination of that information, the utterance of this information and how this utterance is understood or misunderstood [4]. It is not only about what is communicated, so how and how it is grasped by the others.
For a communication at least two subjects are needed and it takes place in a social context. Being social or acting social requires a membership to a kind of society. This attendance at any society takes place through communication and can be accomplished with verbal or non verbal behaviour. So acting unsocial means acting in a way that is against other members of the society or against human being in general. Ignoring other people and do not communicate with them can be an example for such behaviour.

The question „Are we getting unsocial? "in the context of CMC is hard to answer clearly. In public places sometimes people do not interact with the environment, but starring on the screen of their mobile phone (see Figure 1). Even in a personal conversation one of the communication partners looks at times more on the display of his mobile phone than in the eyes of his opponent.
Such behaviour could be described as unsocial. But it is not the technology of CMC itself, but rather the behaviour with the technology in a certain context. The context of using CMC defines if the person acts unsocial or not. So it is a issue of politness in a face-to-face communication and so a problem of society development. The purpose of this paper will not be to question why people communicate more with others via mobile device rather than with their surrounding CMC makes it possible to stay or get in contact with friends, the family or even foreign people alwas and everywhere. Instant Messanger, Social Networks or Dating Sites are only the common ones. In that point of view the unsocial called person acting still social. So being social is a contextual problem
The purpose of this research is to give an insight in CMC. The traits, main types, the possibilities and social effects. Today's CMC has limits to communicate feelings, emotions etc. but there are approaches attempting to overcome these limits.

COMPUTER MEDIATED COMMUNICATION

Decembe et al. described Computer Mediated Communication as „the process by which people create, exchange, and perceive information using networked telecommunications systems (or non-networked computers) that facilitate encoding, transmitting, and decoding "[5]. So CMC describes systems that provides communication or transfering information via computer technology.
In comparison to a natural face-to-face communication there are some basic traits of a computer mediated communication. CMC provides distantless communication for people having access to telecommunication and information technologies (internet, mobilephones etc.). Asynchronous and synchronous messages could be stored and retrieved multiple times and be saved for years. People are able to adress a great number of others in a easy way.
Beside these traits CMC popular techniques can be generally classified into asynchronous and synchronous communication techniques [6]. In the following sections some representatives of each type will be described shortly and some social effects of these CMC methods are listed. Afterwards approaches to enrich the communications channel for emotions, feelings and intimacy are presented.

Asynchronous Communication

Asynchronous CMC is produced when the exchange of information is not simultaneous. It offers a communication independend of time.
E-Mail is the most common and traditional kind of CMC. Its like a letter but with the digital benefits like editing, formating and storage of messages. Big advantages are the independence of distance. People all over the world can send and receive an E-Mail when they have an E-Mail account. Forecasts assume that in 2015 the number of daily send E-Mails is about 204 Billions [7].
Another possibility to communicate with a lot of people is an internet forum. It is a website where users can communicate with posted messages. Those could be read and comment by every registered member. So it offers a place where discussions about certain topics(Threads) or answers could be found. Anonymity is an advantage as well as a disadvantage. On the one hand shy people could dare to participate at discussions and say their opinion. In a face-to-face communication they could be inhibited. On the other hand insulting between members could occur and destroy a discussion. The suspension to say something bad could be reduced.
Usenet is similar to a forum. It is a textbased service for maintaining discussion groups over the internet. Every user need a Newsclient that enables writing and reading messages in the newsgroups. Each topic for discussion is possible and every user can make a comment.
Blogs are websites with information given of one or more persons. Its like a diary about a certain topic or even about the life of a person. Photoblogs, Corporate Blogs or Videoblogs are some examples for possible content. News, informations, opinions and experince could be presented and be commented by a large number of people. „Twitter "is a famous example of this kind of CMC.
In a natural communication the spoken words are only available in one moment. The great advantage of asynchronous communication is the independence of time. That means people have a greater flexibilty in receiving information and responding to messages than in a face-to-face conversation. People have more to think about the information and the respond A conversation takes place in different point of times.

Synchronous Communication

Synchronous CMC is produced when communication occurs simultaneously like in a phonecall or face-to-face conversation. This communication over the internet is called a chat. Differnt forms are classified into textbased and video- and audiobased.

Textbased Chat

The most common textbased methods are Instant Messenging (IM) and Internet Relay Chat (IRC). This types offers realtime communication between two or more people using a client/server architecture. Short textmessages are send via a client and a server stores all user data and messages.
With IRC a lot of people all over the world can be connected at one time and discuss any topics in realtime and create a kind of global group conversation. IM is more private than an IRC and the users usually know each others personally. Usually it is used to have a private conversation with two or more people. Examples for Services using this method are Snapchat, FacebookChat, WhatsApp, Google Talk, Skype, Telegram Messenger, Windows Live Messenger etc.
Textchats are a great way to communicate everywhere and everytime. A kind of emotions could be expressend with symbols called emoticons. But all time availability could create social pressure as well. People claim a immediate

response based on the online status.

Audio and Video Chat

The other forms of a chat use the audio and video channel for the synchronous communication.
Voice over IP (VoIP) is a phonecall in a network or over internet. The costs are often lower than a phonecall over a telecommunication provider. A Videochat or a videoconference is a VoIP with video of the conversation partner. Audio and video data are recorded by mircophone and webcam and transmitted via computer and networks. Mimics and gestures be transmitted this way and has traits of a face-to-face conversation.
Videochat or videoconferencing systems are often used in buisness but used by people in private context as well to stay in contact with relatives over distance [8]. One of the most common service provides is Skype. It combines VoIP, Instant Messenging, videochat and data transfer.
These CMC methods supports social behaviour like in a face-to-face conversation by enabling to see and to hear each other. Sending or receiving information instantly is the main advantage of a synchronous computer mediated communication. It is similar to meeting and talking in real life and presents a more natural way of CMC. But in current Videochat systems like Skype the dialogue partner are more like „Floating Heads "and the oportunity to touch and feel the other person is missing.

Social effects of Computer Mediated Communiation

Computer mediated communication establishs more ways to interact with people. Just 100 years ago the social environment of a person was limited in possibilities to get and stay in contact with friends, familiy members and human environment over distance. There was several commmunication methods like newspaper, telephones and television. The last two were only accessible for richer people, because technique was still expensive. The rest of the society was limited in their personal environment to communicate with other people.
In the last decades technique became cheaper and more people can afford a computer or a mobile phone that enables cheap and easy commuication with people in distance. The capabilities of asynchronous und synchronous CMC had social impacts that were investigated in a wide range and different context [9, 10].
In general people have more freedom to decide when and where they communicate with each other. A face-to-face conversation has its limits in being at the same place. CMC overcomes these limits. E-Mail, IM, SMS or videoconference enables to communicate with family members from everywhere and anytime. Couples can stay in contact in long distance relationships (LDRs) [8]. In a world with many opportunities for travelling or working in foreign countries, people can still communicate with their family and friends. So the personal range for commnication extended throuh CMC.
Computer mediated Communication can help people having difficulties in social interaction. People with autism have often problems with interaction with others. CMC supports partial overcoming this and provide connection with other people and ground their interaction in shared interests [11]. The same applies for deaf people.
Researches reveal that social networks like Facebook are mainly used to stay in contact with old friends [12]. Users with low self esteem are able to accumulate social capital with maintaining large, diffuse networks of friends via Facebook [13]. For lonely people the internet represent a alternative for a real social world. In comparison to non lonely people they used it more for emotional support or to contact others with similar interests [14].
Mobile Devices and the internet connection enable easy communication at any time and place. But this could prompt people to expect immediate response on their message and can cause individual social pressure [9].
Textbased communication methods give the people a sense of anonymity. Influencing factors for a conversation like personal appearance, sympathy, mimics, gestures or the form of articulation are unattended. In social networks, internet communities or Online Games, people can create a digital identity that can differ from their real one [15, e.g.].
Massively multiplayer online role-playing games (MMORPG) build a digital society of people with same interests. Some members replace a real social environment with a digital one and could end even in an addiction [16].
Facebook, Twitter, Instagramm, WhatsApp etc. provide multi media information exchange by messages, pictures, audio or video. These methods enable that a great group of foreign people could get in contact, organize events together or share personal experience with each other. It became easier than with a face-to-face communication or even possible .
The ways how and why CMC has social effects are still a great field of research.

ENRICH COMPUTER MEDIATED COMMUNICATION

Current computer mediated communication methods offer the people to communicate with other people independend from factors like time and distance. These methods were not designed to transfer emotions or feelings but rather to deliver just informations.
A computer mediated conversation between two people is still almost non tangible. In comparison to a face-to-face conversation a textchat is very impersonal. Emoticons and the writing style are sufficient ways to signal emotions or feelings. Phone calls and Videochats give a greater range for emotions via the voice level and volume, mimics and gestures. The next step could be to extend the possibilites in a computer mediated communication by a tangible channel. Make the communication over distance graspable and addressing more sense could support human social activities over network [17]. The people should be allowed to really feel each other.
Following researches investigate how current methods of CMC can be augmented primarily with a haptic channel.

Enrich phone calls

Phone calls are one of the traditional communication ways over distance. They are simliar to a face-to-face conversation but without the visual feedback. Mobile phones provide an easy location-independent communication with other people. In intimate situations it would be appropriate to stroke or just

touch the other person. Simulating the touch of a hand with a vibrotactile feedback is often used as a research approach.
ComTouch is a vibrotactile device sleeve that supports voice communication with touch (Figure 2,b) [18]. The pressure of the hand is translated into a vibrotactile feedback on the other device. It can be used to emphasize the speech, taking turn in a conversation or as a kind of Morse code communication.

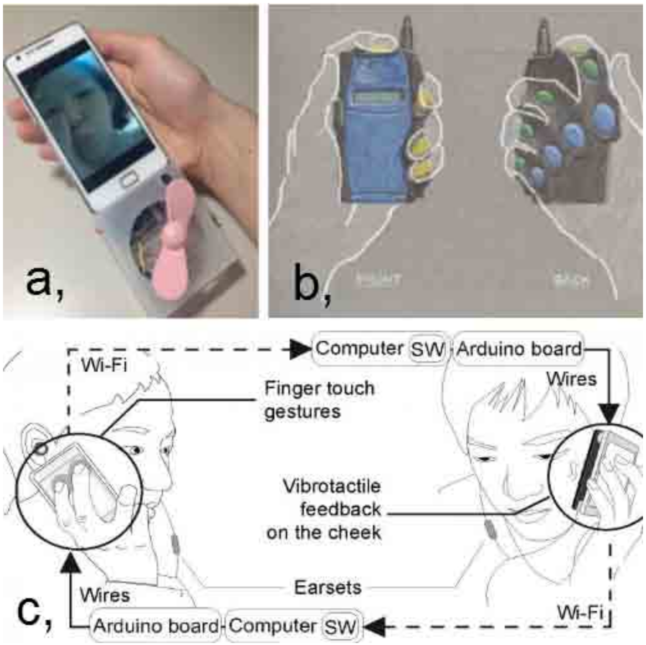


Figure 2. Approaches for tangible communication during a phone call via air (a) or vibrations (b,c)

Using vibrations to transfer greetings, telepresence and emotions are realized with the idea of „pressages". These pressure/vibrotactile messages are provided by ForcePhone, a mobile device system with pressure sensible input and vibrotactile output [19]. Intensity of vibrations on one device depends of the power of squeezing the other device. The study showed non verbal cues are good for emphasizing speech, expressing affection and presence and playfully surprising each other.
CheekTouch combines a tactile feedback and a multifinger input while speaking on the mobile phone (Figure 2,c) [20] . Touching pattern on one phone are converted into vibrotactile patterns on the other phone. Touch behaviours like pinching, stroking, patting, slapping, kissing and tickling can be simulated on the cheek while speaking.
Another mobile approach uses the sense of an airflow to arouse the impression of a face-to-face situation. BlowU is a small fan attached to the bottom of each mobile phone to give a tangible feedback by air (Figure 2,a) [21]. During a video call the gesture of blowing against the screen is recognized from camera and micorphone data by the software. This input is transmitted as a tactile air output on the remote user's mobile. Being close to the face of the other could be illusionated in that way.
A phonecall is a communication method based on speech. There are no ways to see or feel the dialogue partner. Exchanging impersonal information in public by speech could If only impersonal informations are exchanged there is almost no personal problems to speak in public, but talking about private content would be not appropriate for someone.
In puplic places saying lovely words over the mobile phone could be embarrassing for some people. Vibrotactile interaction provides a subtle way to signal such emotions. Sending a private pattern of vibrations can have intimate meanings and wouldn't be noticed rather than spoken words.
Tactile feedback in phone calls offers ways to transfer emotions and extend the options for a social interaction by phone.

Enrich Videochats

Videocalls are a popular multi channel way to communicate with people over distance. That means that the dialogue partner are able to see and hear each other. This CMC comes really close to a face-to-face conversation. The people can see the reactions of the other person by mimics, gestures or body language.
The following approaches try to extend the richness of a videochat with a tactile channel.
Follmer et. al. build a shape changing User Interface. The background was the idea to make digital content graspable [17]. It is a surface consisting of 900 mechanical actuators which can create a 2.5D shape display and are ordered in a grid of 30x30 pins. Each pin can be moved vertically through push-Pull rods connected to computer controlled actuators. Simple physical objects can be displayed as a 2.5D model on the surface. Tracking of users hand interaction and physical objects is realized with an overhead depth camera. An overhead projector provides visual feedback on the surface. The system enables dynamic affordance, constraints and actuation of passive objects [22] (see Figure 4).

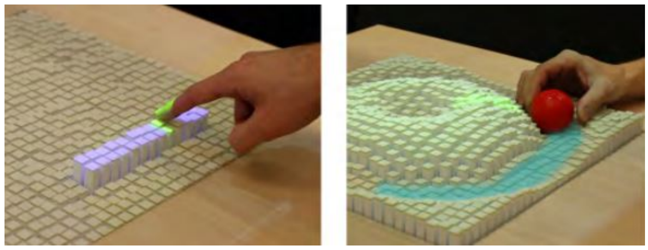


Figure 3. The inForm system provides new interaction techniques for shape changing UIs

Leithinger et. al. used this system to propose the idea of working together remote on physical objects and share digital content. He investigates how people can work collaboratively together over distance and introduces the concept of physical telepresence. Two people can work together on a physical shaped 3D object. The user hand is tracked with a depth camera and represented as a 2.5D model on the remote connected surface of the other user. Cameras and vertical displays provide a visual feedback for the opponent user (see Figure 4) [23].

This set up can be used as a foundation for a tangible interaction during videochats. The shape changing surface provide a way to feel or touch the movement of remote hand/body. It expands the possibilities of existing rich CMC method by a

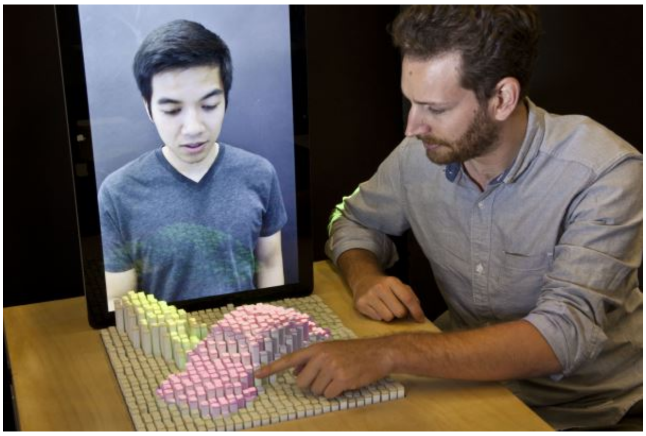


Figure 4. The shape changing surface enables physical representation of 3D objects (red) and the remote hand of the other user (yellow). People could work together or touch indirectly each other.

tactile channel.
In LDRs the inForm system can be used to enrich couples intimate time in videochats. People could play a game with each other (e.g. Balancing a ball togehter) or just share physical presence by touching the hand representation of the partner.
A study revealed that couples want to kiss or to hug the other over the video link. Because of the lack of current possibilities to transfer such interactions in videochats participants develop special gestures. „Every now and then like every night we will typically sign off by giving each other kisses, which is actually like kissing the webcam so it actually looks like you are kissing...“[8].

Figure 5. A couple could transfer a tangible metaphor for a kiss during a videchat with Kissenger

Instead of kissing or hugging the screen or device, transfering the kiss as a tangible feedback to the other person could be an attempt to compensate this lack.
Samani et. al. build a device to transfer a kiss over distance. This physical interface can upgrade remote communication methods like videochats with a methaphor of a kiss. Each user has a remotly connected device which has force sensitive resistors formed like lips. Sensors register the kiss motion and the remote connected actuators transfer this motion into a tangible kiss feeling [24] (see Figure 5).
inForm and Kissenger are two examples for possible ways how current videochats could be enriched with a haptic feedback. These set ups provide visual, aural and tangible interaction methods and show that combining different CMC techniques will increase the options of social interaction over distance.

Emotional CMC with physical artefacts

A private or intimate situation can be characterized that two people are very close together. For couples in Long Distance Relationships sharing physical presence might be a problem. There are different approaches investigate a way to communicate without usual CMC methods. They try to create a feeling of presence of the distant person. The base of these ideas is a remote connected phyiscal object that each person has in their environment. Light or vibrations are possible ways to signal the availablity of one person.
LumiTouch provides active and passive communication between two users via two remotly connected pictures frames [25]. The frames are augmented with touch sensors and LED lights. If one frame is touched the other frame changes its colour as well and signals the presence of the partner. It supports a realtime communication with light and touch pattern to develop a personal emotional language. Glowing pattern can be used to signal a current emotional feeling. A similar but simpler way was the methodology of a bidirectional I/O device, called Feellight [26].
Kowalski et. al. introduces an approach combining linked objects, common communication channels and mobility to enable couples to stay in touch or sharing emotions over distance. Two remote cubes support visual, thermal and tactile feedback and has three message types. An emotional ping (light signal), tap patterns (vibrotactile feedback) and Holding hands (light and thermal feddback). People interact by touching the device. The different feedbacks give a feeling of emotional communication and closeness [27].
The presented approaches are examples how people could signal their presence to each other. Partner could indicate being at home and in the mood to talk. It is like a sign for being ready for communication.
An alternative approach tries to create a real physical link between two seperated users with motion. Brave et al. build a device that provides the illusion that two people can physically interact with each other over distance. inTouch is a shared physical object, that can be manipulated by touching built-in rollers (see Figure 6) [28]. The connected objects are affected from the users manipulation. When one of the rollers is rotated, the corresponding roller on the remote object rotates in the same way. Hand motions and presence of the other person can be transferd and supports a higher level of an interpersonal communication.

Participants announced without visual feedback the interaction felt a little bit awkward. The reason was the insecurity to know with whom the interaction took place. Adding a videochat could be an option to solve that problem [29].

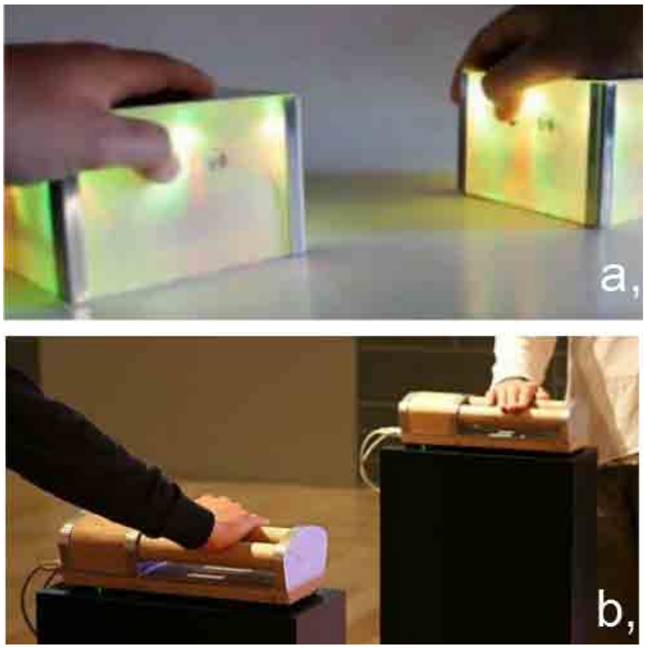


Figure 6. a) The users can indirectly interact with the other user via cubble. b) Rotating cylinders provide a haptic communcation method (inTouch)

Generate the feeling of closeness and intimacy between to people over distance is explored in a wide field of research. Some approaches investigate the idea of remote hugging [30, 31] or being tangible connected through a wearable device like a bracelet [32, 33]. Even whole ambient environments are created which can be used to simulate physical closeness between two persons [34, 35].
These are only a few research approaches representing the different attempts to enrich a personal social interaction over distance. That unusual techniques make the remote interaction more social by generating haptic feedback, emotions, physical closeness and initmacy between seperated people.

DISCUSSION

This paper creates a sense for the impact that CMC has on human communication and social behaviour. Asynchronous and synchronous communication have advantages, disadvantages and limitations in practice in comparison to a traditional face-to-face conversation. Independence of time and distance is the most important improvement. Interpersonal communication has many levels concering the human senses. For emotions or a feeling of closeness and intimacy, CMC has still a lack towards a face-to-face communication. Researches were presented to overcome these lacks. Computer mediated communication supports or even enhance the basic requirement of social interaction to gain new options for a richer interpersonal communication between humans over distance.
Relating to the topic of this paper the question „Are we getting unsocial? "can be cleary denied. Computer Meditated Communication enriched the social communication of people in personal or society. For traditional social interaciton people have to meet other poeple. CMC enables being social even when people stay at home and communicate over networks.
One person could adress more other people easier than by a face-to-face conversation. People chat with each other by IM or plan their leasure time activities in social networks. Communication with friends, the family or foreign people everywhere and everytime is supported through the development of the internet, computers technology, mobile devices and software services. Smartphones offer on the way communication with other people. But the possibility being online and responding immediately could create personal pressure as well. Computer mediated communication bridges geographical distance between two people and provide staying in contact. Couples in LDRs use CMC to talk and to see each other or just to create a connection between the seperated personal environment. But there is still the problem of transferring physical interaction. Researches investigate different ways to transfer touching and feeling the other person. Vibrations, light, motions and even thermal signals were transmitted to create the feeling of physical closeness.
Beside these new possibilites of communication the old values of respect between humans has to be mentioned. During a face-to-face conversation it is unpolite doing something else. Interacting with the mobile phone to communicate with other although sitting at a table with the family is disrespectful. The attention is not on the current social activity, but somewhere else. In this case the person has an unsocial behaviour in an actual social situation (social marginalization).
Computer Mediated communication is enhanced in a wide range to improve interpersonal social interaction. But handling with this technology in social situations has to be developed and adjusted as well. A first step to comply with rules of conduct of CMC usage in face-to-face conversations could be a kind of etiquette guide (e.g. current Netiquette in internet foren).

CONCLUSION

After a short definition of communication and social behaviour this paper reviewed some basics of computer mediated communication, the definition, the main traits and some general types. Examples for social effect of such methods on human behaviour and social interaction were presented.
Computer Mediated Communication increases the possibilities to communicate with each other. Great advantages are the independence of time and an easy communication over distance with almost every person on the world. But there is still a lack of transfering haptic interpersonal interaction to create emotional closeness and a kind of intimacy. Several approachs to upgrade the visual and auditive channel in differnt ways were presented.
Relating to the topic of this paper computer mediated communcation offers a wide range of opportunities for acting social. But being social depends on the context when CMC is used. Chatting with others in a face-to-face conversation is unpolite or unsocial just as reading a book. So not the technology is the reason for unsocial interaction it is the dealing with it. Social rules in dealing with CMC in face-to-face situation has to be develop and be socially accepted.

REFERENCES

1. 2013. http://stopphubbing.com/.
2. 2014. http://www.emarketer.com/Article/Youre-Map-Happy-You-Know-Tap-Your-Screen/1011077.
3. http://www.thefreedictionary.com/communication.
4. Niklas Luhmann. What is communication? *Communication theory*, 2(3):251–259, 1992.
5. J. December. What is computer-mediated communication?, 1996. http://www.december.com/john/study/cmc/what.html.
6. Alan Dix. *Human-computer interaction*. Springer, 2009.
7. 2015. http://de.statista.com.
8. Carman Neustaedter and Saul Greenberg. Intimacy in long-distance relationships over video chat. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 753–762. ACM, 2012.
9. Sara Kiesler, Jane Siegel, and Timothy W McGuire. Social psychological aspects of computer-mediated communication. *American psychologist*, 39(10):1123, 1984.
10. Joseph B Walther. Computer-mediated communication impersonal, interpersonal, and hyperpersonal interaction. *Communication research*, 23(1):3–43, 1996.
11. Moira Burke, Robert Kraut, and Diane Williams. Social use of computer-mediated communication by adults on the autism spectrum. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pages 425–434. ACM, 2010.
12. Nicole B Ellison, Charles Steinfield, and Cliff Lampe. The benefits of facebook friends: social capital and college students use of online social network sites. *Journal of Computer-Mediated Communication*, 12(4):1143–1168, 2007.
13. Charles Steinfield, Nicole B Ellison, and Cliff Lampe. Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. *Journal of Applied Developmental Psychology*, 29(6):434–445, 2008.
14. Janet Morahan-Martin and Phyllis Schumacher. Loneliness and social uses of the internet. *Computers in Human Behavior*, 19(6):659–671, 2003.
15. Shanyang Zhao, Sherri Grasmuck, and Jason Martin. Identity construction on facebook: Digital empowerment in anchored relationships. *Computers in human behavior*, 24(5):1816–1836, 2008.
16. Daria Joanna Kuss and Mark D Griffiths. Internet gaming addiction: A systematic review of empirical research. *International Journal of Mental Health and Addiction*, 10(2):278–296, 2012.
17. Hiroshi Ishii and Brygg Ullmer. Tangible bits: towards seamless interfaces between people, bits and atoms. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, pages 234–241. ACM, 1997.
18. Angela Chang, Sile O'Modhrain, Rob Jacob, Eric Gunther, and Hiroshi Ishii. Comtouch: design of a vibrotactile communication device. In *Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques*, pages 312–320. ACM, 2002.
19. Eve Hoggan, Craig Stewart, Laura Haverinen, Giulio Jacucci, and Vuokko Lantz. Pressages: augmenting phone calls with non-verbal messages. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, pages 555–562. ACM, 2012.
20. Young-Woo Park, Chang-Young Lim, and Tek-Jin Nam. Cheektouch: an affective interaction technique while speaking on the mobile phone. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, pages 3241–3246. ACM, 2010.
21. Inkyung Choi. Blowu: physical feedback for seamless remote interaction in mobile. In *SIGGRAPH Asia 2014 Emerging Technologies*, page 2. ACM, 2014.
22. Sean Follmer, Daniel Leithinger, Alex Olwal, Akimitsu Hogge, and Hiroshi Ishii. inform: dynamic physical affordances and constraints through shape and object actuation. In *UIST*, pages 417–426, 2013.
23. Daniel Leithinger, Sean Follmer, Alex Olwal, and Hiroshi Ishii. Physical telepresence: shape capture and display for embodied, computer-mediated remote collaboration. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pages 461–470. ACM, 2014.
24. Hooman Aghaebrahimi Samani, Rahul Parsani, Lenis Tejada Rodriguez, Elham Saadatian, Kumudu Harshadeva Dissanayake, and Adrian David Cheok. Kissenger: design of a kiss transmission device. In *Proceedings of the Designing Interactive Systems Conference*, pages 48–57. ACM, 2012.
25. Angela Chang, Ben Resner, Brad Koerner, XingChen Wang, and Hiroshi Ishii. Lumitouch: an emotional communication device. In *CHI'01 extended abstracts on Human factors in computing systems*, pages 313–314. ACM, 2001.
26. Kenji Suzuki and Shuji Hashimoto. Feellight: a communication device for distant nonverbal exchange. In *Proceedings of the 2004 ACM SIGMM workshop on Effective telepresence*, pages 40–44. ACM, 2004.
27. Robert Kowalski, Sebastian Loehmann, and Doris Hausen. cubble: a multi-device hybrid approach supporting communication in long-distance relationships. In *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction*, pages 201–204. ACM, 2013.

28. Scott Brave and Andrew Dahley. intouch: a medium for haptic interpersonal communication. In *CHI'97 Extended Abstracts on Human Factors in Computing Systems*, pages 363–364. ACM, 1997.
29. Scott Brave, Hiroshi Ishii, and Andrew Dahley. Tangible interfaces for remote collaboration and communication. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, pages 169–178. ACM, 1998.
30. Florian'Floyd' Mueller, Frank Vetere, Martin R Gibbs, Jesper Kjeldskov, Sonja Pedell, and Steve Howard. Hug over a distance. In *CHI'05 extended abstracts on Human factors in computing systems*, pages 1673–1676. ACM, 2005.
31. Carl DiSalvo, Francine Gemperle, Jodi Forlizzi, and Elliott Montgomery. The hug: an exploration of robotic form for intimate communication. In *Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003. The 12th IEEE International Workshop on*, pages 403–408. IEEE, 2003.
32. 2013. http://tactilu.com/.
33. Minna Pakanen, Ashley Colley, Jonna Häkkilä, Johan Kildal, and Vuokko Lantz. Squeezy bracelet: designing a wearable communication device for tactile interaction. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*, pages 305–314. ACM, 2014.
34. Hiroshi Ishii, Craig Wisneski, Scott Brave, Andrew Dahley, Matt Gorbet, Brygg Ullmer, and Paul Yarin. ambientroom: integrating ambient media with architectural space. In *CHI 98 Cconference Summary on Human Factors in Computing Systems*, pages 173–174. ACM, 1998.
35. Chris Dodge. The bed: a medium for intimate communication. In *CHI'97 Extended Abstracts on Human Factors in Computing Systems*, pages 371–372. ACM, 1997.

Interaction Concepts for People with Disabilities

Andrea Reidel
Institute of Media Informatics
Ulm, University
andrea.reidel@uni-ulm.de

ABSTRACT
This paper gives an overview of some concepts researchers developed for people with disabilities, so they can interact with technologies like smartphones, tablets, the Internet, self-driving cars and some more. Below, there are listed the requirements such interaction concepts have to fulfill. Furthermore, there will be an evaluation of these ideas to inspect if they really abolish the barriers between disabled people and contemporary technologies.

Author Keywords
Interaction concepts; Accessibility; Visual impairment; Cognitive disability; Motor disability.

INTRODUCTION

To most people it is self-evident to have access to contemporary technologies like smartphones or computers, the Internet, but also cars or other means of transportation. But there is a part of society suffering from disabilities that make access to these technologies difficult. According to [15] about 15% of the world's population are afflicted with serious physical or cognitive disabilities and about 20% of them are living in developed countries. It follows that one out of five persons encounter difficulties pertaining to accessibility of modern interactive systems.

For several years now, researchers concentrate on developing interaction concepts to abolish the barriers between disabled people and the latest technologies. Below, there will be presented a couple of the latest interaction techniques for disabled people, primarily for visual, cognitive and motor impaired ones.

REQUIREMENTS

As disabled people are limited in several cognitive or physical ways, supporting technologies have to fulfill some demands.

Visual impaired persons are limited in their everyday life, because they lost a very important sense: their sight. That makes e.g. accessibility to the web, the 'normal' use of a mobile phone or a PC and even moving around difficult. Besides, most blind people strongly depend on the support of others. In the long term, this dependency and helplessness can lead to depressions, less self-esteem and social isolation [18, 12]. To offer visual impaired people access to the Internet, it is important to support them by using their well functioning senses like hearing or touching. So, other devices like mobile phones, navigation systems, gaming consoles, tablets or cars also have to address these other senses and provide non-visual feedback. In the two papers [9] and [12] researchers concerned themselves with the navigation of blind people inside and outside their houses. Below, there will be presented concepts that have potential to raise blind persons' mobility and makes them more independent, offer them access to pictures on the web and provide tactile feedback on a tablet.

People with cognitive disabilities have to deal with some other problems like paying attention or concentrating on something. In some cases those complications are only temporary, so the persons concerned can be healed with a therapy. Therefore, new technologies like tabletop systems can be used for the healing process. Is a cognitive impairment permanent, people concerned could be offered e.g. entertainment.

Motor impaired persons are suffering from physical problems like the inability to make precise movements with their arms. The outcome of this is that motor disabled people cannot hit small buttons with a mouse [21] or make exact touch gestures to control an application.

Finally, researchers have to deal with the specific difficulties their target audience implicates and involve them in the development process. This is quite important, because disabled people also should have access to new technologies, just as healthy ones. Also a noteworthy demand disabled people have, is access to entertainment like games on their PCs, smartphones or tablets. There is already a wide range of applications, but most of them do not offer the relevant usability for disabled user.

INTERACTION CONCEPTS

Visual impairments

There are already existing some interaction concepts for visual impaired people open to the public, like Braille displays [2] and screen readers [5], but researchers are constantly searching for further and maybe more effective and convenient concepts. Below, there will be presented some novel approaches in the range of interaction concepts, especially for visual impaired people.

The blind driver challenge

The topic of self-driving cars came to the fore in 2004, when the National Federation of the Blind appealed researchers for developing a non-visual interface. It should provide

7th Seminar on Research Trends in Media Informatics (RTMI '15). February 2015, Ulm University, Ulm, Germany.

detailed information about the car's environment [4]. Due to this challenge Google produced an autonomous car [14]. Google's researchers used an Toyota Prius model equipped with a 360° multilayer laser scanner and also mechanisms to control the car's travel direction and speed (Figure 1).

Figure 1. Google's self-driving car, a Toyota Prius model [14]

Burkay Sucu and Eelke Folmer also dealt with the topic of a self-driving car that would enable blind people being mobile and move independently [18]. What is special about their approach is the steering mechanism. They designed an interface that helps both blind and sighted drivers to stay in their lane. They mentioned that sighted people also can be temporary affected through glare [17].
The inspiration for the authors' work is the tactile feedback rumble strips or Bott's dots give the car driver if he or she leaves the track. Accordingly the steering wheel has a built-in vibrator on the right and the left side. As shown in Figure 3 the driver moves the wheel to the right if he or she feels a vibration on the left side of the steering wheel. But Burkay Sucu and Eelke Folmer not only intended to inform the user when to steer, but also how far to move the steering wheel. It follows that the vibration does not stop before the wheel is in the right position named T. In [18] they expanded T to an area around the calculated value to reduce the oscillating motion while trying to hold the wheel at position T. The steering interface only guides the driver through a curve, but does not adjust the deviating position of the car while driving on a straight road. But the researchers improved their interface and implemented an self-correction mechanism pictured in Figure 2.

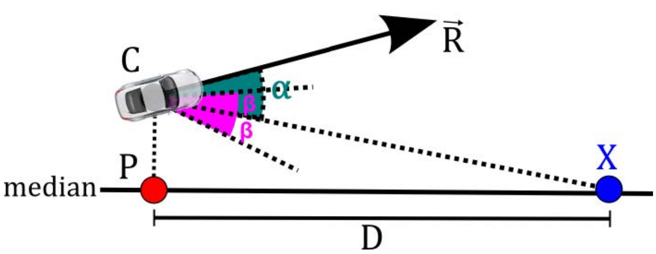


Figure 2. Self-correction mechanism of the steering interface in [18]

Firstly, in [18] they compute the target position X and the median, as well as the car's current position C. With that information the authors receive the spot P on the median and the distance between P and X. Are P or X located on a curve, D is the length of the curve's arc.
Also important is the car's driving direction $\overrightarrow{R}$ and the angle α between $\overrightarrow{CX}$ and $\overrightarrow{R}$. If the car's direction deviates too much from $\overrightarrow{CX}$, at least one bound of α is located outside of the dead-band window β enclosing $\overrightarrow{CX}$. In that case, the steering wheel gives tactile feedback and requests the user to adjust the driving direction.
The authors Burkay Sucu and Eelke Folmer conducted a study to evaluate the use of their interface while driving. They compared two cases. The driver receives either visual and haptic or only haptic feedback. The result showed that the self-correction mechanism really diminishes the deviation from the curve's median. They also measured the reaction time of the participants to the haptic feedback from the steering wheel. According to [18], blind people are more sensitive to haptic feedback than sighted ones. The study results showed a less reaction time for the blind drivers than the value of the group with both visual and haptic feedback. However, they conducted the study just in a simulator and not integrated in an real vehicle driving on streets in the real world. Anyway the researchers got important findings about supporting self-driving cars.

Tactile display using TeslaTouch

To provide tactile feedback for visual impaired people while using a touch screen, Cheng Xu, Ali Israr, Ivan Poupyrev, Olivier Bau and Chris Harrison used 'TeslaTouch' for the realization of their application in [22]. They implemented an interface that enables sensing e.g. dots, braille letters or images on a screen without using mechanical properties.
'TeslaTouch' utilizes electrovibration [8] to let the user perceive what is displayed on the screen. The used touch panel, in this case a 3M MicroTouch Display, consists of a transparent electrode upon a glass plate covered with an isolator layer. The electrode is excited with a periodical electric signal. If the user's finger contacts the top layer and moves it like in Figure 4, there is a voltage difference between the finger and the electrode. Thus, the user feels a higher force of attraction to the display.

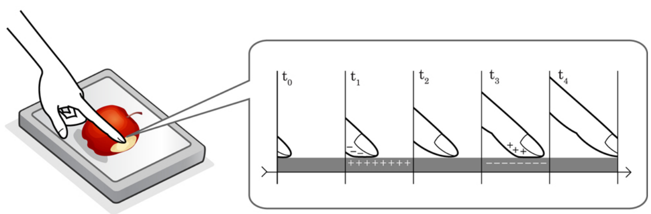


Figure 4. The user can feel textures on the screen, because there is a voltage difference between the finger and the layer [8]

In [22], participants of a carried out study described the feeling of sensing displayed dots like 'sticky' regions, changing friction or a chalk board. To represent braille letters, they had multiple approaches, because one single letter consists of a

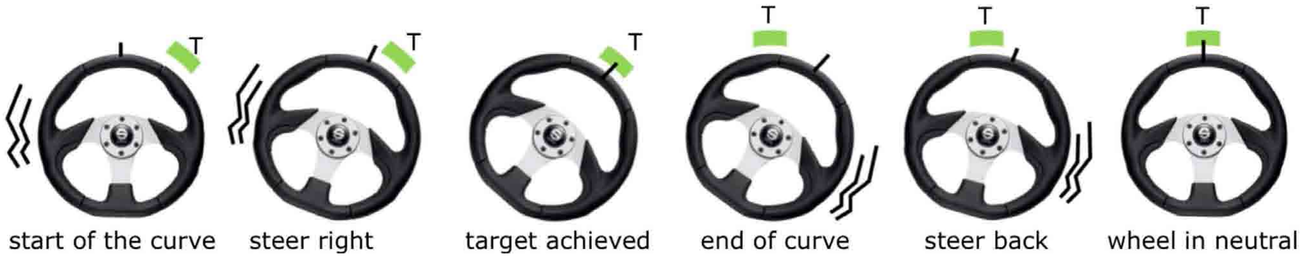


Figure 3. The vibration principle of the steering interface in [18]

defined matrix of six dots. It is a challenge to represent these components well readable for blind people. That is why the researchers implemented the representation of braille letters in several ways. Firstly, with the modulation of the used frequency. Secondly, with temporal mapping, i.e. by and by the individual dots are displayed with a defined interval in between, and at last with spatial separation.

After testing the different techniques, it emerged that none of them is flawless. The results they gained from their test regarding to identify simple geometrical objects on the screen are better. The participants had to identify a circle, a triangle and a square. Each figure was shown in a solid, outline and solid with outline style. According to the findings, the solid shapes were most convenient to identify.

The setup they used for their study looked like depicted in Figure 5. The touch panel sends the finger's location to a connected PC, whereupon the appropriate sinusoidal signal is calculated. Thereby, the researches are able to manipulate the frequency to that effect that the user can sense several textures. The output signal is amplified by the connected TeslaTouch Driver and transmitted to a wristband the user is wearing.

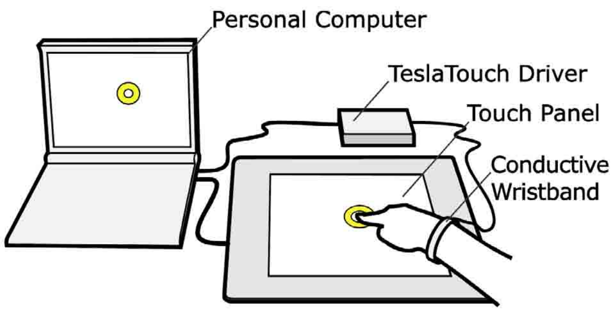


Figure 5. The setup: The touch display using TeslaTouch and the user's wristband are connected to a PC. A driver translates the incoming data [22].

The participants joined the study showed great interesting in the interaction with TeslaTouch, although there were several problems regarding to recognition of displayed content.

Sonification of data

As blind people are not able to use their sight, they have to rely on other senses like touching and hearing. Especially the last one is interesting for some researchers like Haixia Zhao [24]. There are already some text-to-speech solutions like screen readers [5], but Zhao is rather focused on non-speech sounds. There are some other researchers interested in this topic like Micheal Banf and Volker Blanz [7]. Zhao occupied himself with the sonification of data, whereas Banf and Blanz want to make images accessible for visual impaired people.

To sonify pictures, they had to resolve the questions, which information from an image is important and how it could be represented through sound. Their aim is it to sonify different level of an image, e.g. colors, textures like roughness or the detachment between nature and man-made structures. The authors supposed that sonification of pictures could make the Internet more accessible for blind people. The therefore implemented software describes the arising colors with the HSL model (hue h, saturation s and lightness l). They used this model, because Banf and Blanz thought it is a more intuitive color description as the RGB model.

When they analyzed a picture, they faced a problem: some images include huge contrasts from pixel to pixel, because the camera caused image noise. The variable sonification of these regions could be a impertinence for the user. Concerning this finding they used *bilateral filtering* [19] to smooth the concerned regions.

When they further dealt with the used sounds, they decided to realize it like depicted in Figure 6. They thought instruments would be a better representation of objects than colors. The authors concentrated on the opposing color pairs blue-yellow and red-green. These colors are represented through complementary sound characteristics like light wind instruments and bass or constant beats and Tremolo. To be able to make decisions about suitable sounds, both of the authors busied themselves with the color theory, particularly with the separation of cold and warm colors.

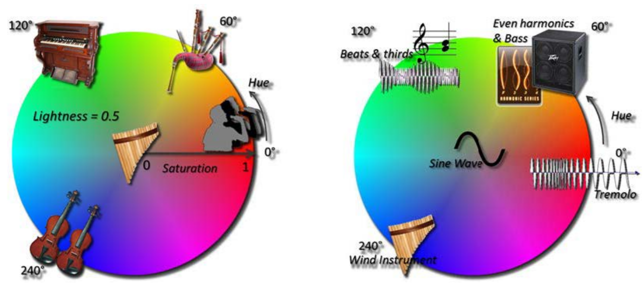


Figure 6. Left: the representation of opponent colors through MIDI instruments they developed in another paper. Right: the new model. Colors are served by Complementary Sound Characteristics. [7]

To serve the lightness of colors, they mapped the value l to the eight notes of the gamut, whereupon a low tone repre-

sents a low lightness value. With that preparation and a self-made synthesis model they were able to represent colors as follows: first, there is a single sine wave and depending on the lightness of the color, the pitch will be adjusted. Then, the appropriate sound is added according to the hue value. As an example, a beat of two frequencies is added if h indicates red. These frequencies are adjusted depending on how intensive the color is.
To afford the user to distinguish between natural regions and buildings, Banf and Blanz [7] grouped the image in areas of 16x16 pixels. For each of these squares a feature vector is calculated on the basis of the corresponding histogram. With that vectors, classifications can be made. Further, they used object detection and recognition algorithms and generated a result like in Figure 7.

Figure 7. Results of the used algorithms to detect nature regions and man-built structures [7]

Haixia Zhao developed an interactive tool called *iSonic* [24]. It enables visual impaired people working on abstract data, like the number of inhabitants of all the states in the USA. The values are available in a table or in form of a map. By using the arrows on the keyboard the user can navigate on the map from state to state and listen to the spoken information. In this context, the interacting person can choose which information are to be read out, like the country's name or the number of population. If the user wants to select some information in the table, e.g. five specific states, they are also selected on the map. The interaction on the map interface is also possible by using the number fields on the keyboard. Thus, the user can pick one of the regions that is mapped to the arrangement of the numbers. That means, area seven is in the top left corner of the map and includes the states that are located in the northwest. To support the user with the navigation on the map, the application gives auditive feedback. A voice tells the user, when borders are crossed or all states are visited once.

Cognitive disabilities

It is important to understand, how cognitive disabled persons behave and what demands they make on technical devices. With that finding it is possible to offer them suitable interaction concepts.

Tabletop interactions for people with cognitive disabilities

Ru Zarin and Daniel Fallman [23] developed a tabletop multitouch system that offers children suffering from Autism Spectrum Disorder (ASD) or Down's Syndrome a way to train their speech and language skills. Just at younger ages it is important to gain and enhance that abilities, because they are essential for further acquirement of skills.
The tabletop multitouch system, called 'Trollskogen', features a number of micro applications for the users. The main screen of the system displays a simple illustrated forest with all micro applications represented as mushrooms. In [23] they implemented four applications.
'The Forest Cabin Program' allows the user to explore a room with three areas and interactive elements within. Supplementary a voice tells a corresponding story about the place and the people living there. Thereby cognitive disabled children can learn how to behave in certain situations like entering a room or having lunch with others.
The second application is called 'The Dancing Troll Program' and enables the user moving a figure by using the voice. By giving the troll character directions the user advances intonation and enunciation.
A further application named 'The Finger Paint Program' stimulates the creativity of the user. It offers a color palette and a couple of figures from the Trollskogen scheme to color. With that it is possible to draw pictures or write on a canvas. This sort of activity could calm a child down.
The last application is called 'Talking Symbols'. In Sweden, where Ru Zarin and Daniel Fallman come from, working with symbols is very popular. Therefore they developed 'Talking Symbols' to help cognitive disabled children forming sentences with pictographs. Are two or more elements brought together, the system connects them to a sentence. Then the user can listen to the built symbol construction by moving a speaker icon to the connected pictographs. While the system reads out the sentence, the suitable symbol is highlighted.
'Trollskogen' is tested with six cognitive impaired children in a group and afterwards individually with their teacher in a classroom. The reaction from both groups was mainly positive. Especially the painting program succeeded, because it amused the children most.

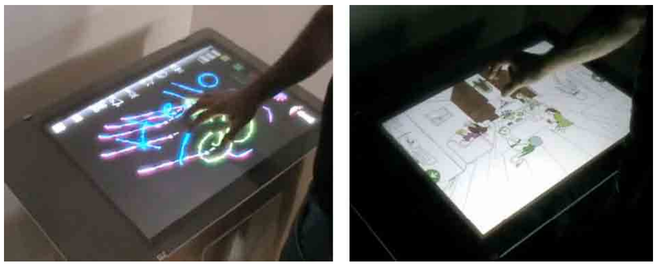


Figure 8. Demonstration of the micro applications 'The Finger Paint Program' and 'The Forest Cabin Program' in [23]

Another interesting interaction concept is the 'fun.tast.tisch'-project by Mirjam Augstein, Thomas Neumayr, Renate Ruckser-Scherb, Isabel Karlhuber and Josef Altmann from Austria [6]. They developed an interactive tabletop system that helps people to rehabilitate after acquiring brain injuries. Conventional therapies implicate some disadvantages like an intricate training setup or the time-consuming generation of process statistics. To improve these and more drawbacks, Augstein *et al* evolved an assistive technology in terms of a multitouch tabletop system.
For cognitive disabled people it is important to regain and train attention, memory and visuo-spatial skills. Therefore,

the 'fun.tast.tisch' offers an application where affected people have to solve puzzles in the form of Tangram, a Chinese puzzle game [20].
The tabletop interface shows a figure, consisting of at most seven parts, the user has to copy with real building blocks called 'tans'. Identifying not only touch gestures but also objects is a distinguished advantage of the used tabletop Samsung SUR40 with Microsoft PixelSense [3].
If the user matches the shown figure with the tans, the tabletop gives the user visual and optional auditive feedback. Because the pieces are transparent, the acknowledgment is quite obvious for the user, although the tans lie above the shown pattern (Figure 9). During the training session the therapist has the possibility to adjust the level of difficulty directly at the tabletop.
The 'fun.tast.tisch' has been tested with six cognitive disabled people and four therapists. Both groups stated that using the tabletop was comfortable and helpful. The study also uncovered some drawbacks, but on the whole the reaction was positive.

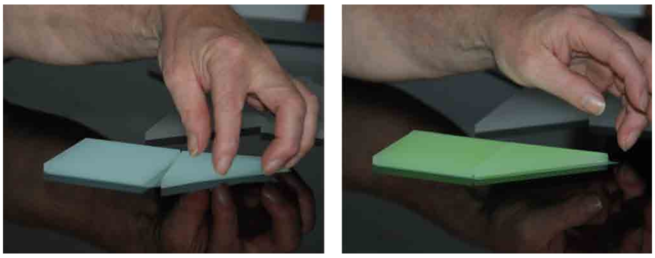


Figure 9. A patient during her training session using the 'fun.tast.tisch'. She finishes an exercise and receives visual feedback [6].

Motor disabilities

There are several types of motor impairments, but in the following chapter, there are mainly persons in the focus that cannot move their arms as a healthy person. If someone cannot use his or her hands to interact, there has to be other ways to make technical devices accessible.

Hands-Free Gesture Control

According to [13], gestures performed with hands are an effective and rife type of interaction. Unfortunately people with motor disabilities are usually not able to use this way of interaction with a technical device effectively. Therefore researchers keep themselves busy with that topic and try to envelope hands-free controllable interfaces. Marco Hirsch *et al* implemented such a system for motor impaired people [13].
One of their inspirations was to protect the user's privacy and avoid the utilization of cameras. To track the movements of a person anyway, the authors use the active capacitive sensing principle. That means, there are four capacitive electrodes integrated into a wearable neckband (Figure 10) that measures changes inside the user's body. The idea has been developed over several years.

In 2010, Jingyuan Cheng, Oliver Amft and Paul Lukowicz published it for the first time [11]. The fundamental thoughts are described there, but there are still problems to solve. One year later, they published the next generation of their interface [10]. It allows recording data with a via Bluetooth connected

Figure 10. The prototype of the capacitive neckband [13].

smartphone. On the basis of these data, the authors could analyze, at what time a person goes for a meal (swallowing) or sleeps (quiet period). They improved their interface, so it can recognize head movements and react to the gestures [13].
The neckband consists of a capacitor, more precisely a dielectric material surrounded by conductive layers. In this case, one conductive plane is located in the neckband, whereas the neck's inside replaces the dielectric material. Thus, by every movement of muscles, tendons or vasculature originates a electric signal, especially if the user moves his or her head. To measure the electric changes correctly, the electrodes have to be placed properly. Like pictured in Figure 11 the arrangement is symmetrical.

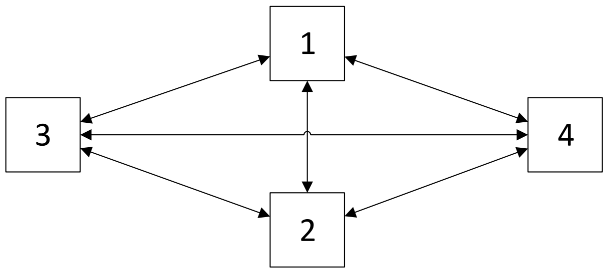


Figure 11. Arrangement of the four electrodes in the neckband.

Two of them are located above and under the larynx, the others are placed on the right and left side of the neck. On the basis of the electrodes, changes can be calculated from the modified difference between the them. The neckband is connected to a circuit board that amplifies and converts the signal to digital data and finally transmits them wireless.
To evaluate their interface, Marco Hirsch *et al* tested it with 12 participants. They prepared an application that gives the user text and voice instructions and requests him or her to perform 15 varied head gestures. This gestures are made up of five basic movements:

- nodding: up and down movement
- tilting: movement to the right or left
- looking to the right or left
- circling the head clockwise or anti-clockwise
- move the head like a woodpecker.

Through combining these movements the researchers created gestures like 'double woodpecker move' or fast and slowly

moves. But they not only concentrated on head gestures, but also on head postures. Thereto the developers prepared 15 designations in form of a 3x5-matrix on a wall. In addition, there were defined four points in extreme angles: above, at the bottom, on the right and on the left side. The participants got instructions via voice. The responsible persons tracks the head movements with a laser pointer placed on the head of the subject.
Marco Hirsch *et al* mapped both studies to possible application scenarios. The first that affects the head gestures could be the control of an electric wheelchair. Below there is one possible mapping for the gestures:

- double nod up to increase the speed
- double nod down to decelerate
- double tilt left to steer to the left
- double tilt right to steer to the right
- double woodpecker move to force an emergency stop.

For the second study they devised the navigation of a photo slide show. In their survey, there were four extreme angles (above, at the bottom, on the right and on the left side) and 35 further spots. To navigate through the pictures, only the furthermost labellings are used both to step forward or backward and jump to the first or last picture in the slide show.

Another paper that deals with the inaccessibility of technologies for motor impaired people is [16]. Adam Sporka, Ben Carson, Paul Nauert and Sri Kurniawan analyze by means of a user study where the difficulties referring to accessibility to composition software for musicians are and how they can be negotiated.
Interviews with three motor impaired persons yielded some suggestions for improvements. Among other things, it should be possible to proceed like you would compose on paper. That means, the composer assesses approximates sections and improves them little by little, or relieves the introduction to such a software. But the most important suggestion is the improvement of the on-screen keyboard, so that motor disabled people can easier enter musical symbols. In this context one of the participants mentioned an advantageous text entry tool named Dasher [1], that supports input of alphabetic characters through navigating with one finger.
In general the topic of software for disabled musicians is apparently an unexplored domain in research, but as well as in other subject areas the people concerned are deserving of get access to more convenient systems.

EVALUATION

Now, it is time to judge how far the introduced concepts fulfill the demands of people with disabilities. All of them offer persons concerned fitting access possibilities on the whole, but in detail there are still some aspects to improve.
First, the techniques for visual impaired people. The self-driving car and the steering interface are important projects to make blind persons independent and raise their self-confidence. The control of those precursors is intuitive and suitable for the daily use. Nevertheless it will take time, until blind people really drive alone. The laws in some countries forbid them to drive on their own and enjoin them to have a companion. But while the technology is going to be improved and more safe, it is possible that it comes to rethinking of this topic in society.
The steering interface is a really promising concept and a step into independence for visual impaired people. By implementing the self-correction mechanism, this concepts got more safely for drivers, because also outside of curves there is monitored that the car does not leave the track. The study, the authors made, showed a average standard deviation of 0.97 meter and therefore a value that hypothetical enables safe driving. Since the haptic interface is not tested in a real vehicle, it is not possible to say if it really acts the same way in reality.
TeslaTouch is also a promising idea, because it offers visual impaired persons access to content showed on a touch display. Interacting with the touch panel is intuitive and easy. Additionally, the realization of these concept is not very expensive. That is an important advice, because people concerned have to be able to afford it.
The carried out study has shown problems according to recognizing braille letters or thin lines. The components are that small, the users cannot recognize them with their fingers easily. Although they have realized three different ways of presenting braille letters, none of them was ideal and convenient for the participants. Due to the rather poor sensitivity of fingertips, geometrical shapes cannot displayed with outline style as well. However, TeslaTouch can be used to access other information that is not as fine as little braille dots.
Maybe the concept of sonification of data or images is not attractive for everyone, especially the mapping between pictures and the chosen sounds. But particularly, in this case subjects from the target audience have to be involved. Someone who is sighted perhaps cannot understand how people that have been always blind, experience colors. On the whole, sonification is a great concept to make, above all huge numerical data, accessible without seeing them.
The two tabletop applications for cognitive disabled people are a valuable help. Touch interaction is, as said above, an intuitive way to communicate with a technical device. In matters of cognitive disabled people, maybe this advantage is more important anyway. The authors of the 'fun.tast.tisch' project achieved their goal and improved the used objects, display motivating feedback for the users, make changing the degree of difficulty easier and above all, enable the simple generation of statistics. That makes the 'fun.tast.tisch' to a really helpful support in therapies. There are also some little aspects that could be advanced, but generally speaking this works fine.
The tabletop for the cognitive disabled children is also a good supporting system, because as a study with six children showed, they like the interaction with the touch panel. If children think something is interesting, it is supposable that their learning progress grows better. The teacher has the possibility to easily swap the applications if he or she notices the child is bored.
To build a special interaction possibility for motor impaired people is not easy, but the neckband that recognizes head ges-

tures and postures is quite interesting. It is imaginable that it is a little bit uncomfortable, to wear something around the neck. Even if the user can hide it under a pullover. If a user accepts these circumstances, the neckband is quite useful. The mapping to a wheelchair control is awesome, provided, it interprets uncontrolled moves not as a command, but if the security of these concept is improved enough, it would be an relative intuitive control.
Generally speaking, all of the introduced concepts work, but there are still some incongruities.

CONCLUSION AND FUTURE WORK

There are lots of interesting interaction concepts researchers developed for disabled people over the last years. Since it is normal to have a smartphone or a tablet, the solutions researchers develop could be used with these devices. The gadgets named above can be involved to support ideas like collecting and saving data from the neckband in [13]. Or in case of an emergency, a smartphone could send a message or make a call
Many ideas have potential, but perhaps getting accessible for the people concerned will take a long time, because the concepts are not totally mature or not evaluated enough. It is quite important continuing to involve affected persons in the development processes, because they know best, weather an idea works or not. It is really necessary to search for new finding and improve existing techniques, so that disabled people have the chance to use contemporary technologies as well as healthy persons. Therefore, the fields of sonification and haptic feedback should be explored forwards. Techniques like TeslaTouch are a really useful way of providing tactile feedback for disabled users. Sonification stimulates fantasy and creativity of visual impaired user and of course makes information accessible.

REFERENCES

1. Demonstration of dasher, a text entry tool. `https://www.youtube.com/watch?v=0d6yIquOKQ0`. Accessed: 2014-12-21.
2. Deutscher hilfsmittelversand: Braillezeilen. `http://www.deutscher-hilfsmittelversand.de/produkte/brz/index.html`. Accessed: 2014-12-12.
3. Microsoft pixelsense. `http://www.microsoft.com/en-us/pixelsense/default.aspx`. Accessed: 2014-12-13.
4. National federation of the blind: Blind driver challenge. `http://blinddriverchallenge.org/about-the-blind-driver-challenge`. Accessed: 2014-12-18.
5. Nv access: Home of the free nvda screen reader. `http://www.nvaccess.org`. Accessed: 2014-12-14.
6. Augstein, M., Neumayr, T., Ruckser-Scherb, R., Karlhuber, I., and Altmann, J. The fun. tast. tisch. project: a novel approach to neuro-rehabilitation using an interactive multiuser multitouch tabletop. In *Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces*, ACM (2013), 81–90.
7. Banf, M., and Blanz, V. Sonification of images for the visually impaired using a multi-level approach. In *Proceedings of the 4th Augmented Human International Conference*, ACM (2013), 162–169.
8. Bau, O., Poupyrev, I., Israr, A., and Harrison, C. Teslatouch: Electrovibration for touch surfaces. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, ACM (New York, NY, USA, 2010), 283–292.
9. Brock, M., and Kristensson, P. O. 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, 2013), 255–258.
10. Cheng, J., Amft, O., and Lukowicz, P. Active capacitive sensing: Exploring a new wearable sensing modality for activity recognition. In *Pervasive Computing*. Springer, 2010, 319–336.
11. Cheng, J., Zhou, B., Kunze, K., Rheinländer, C. C., Wille, S., Wehn, N., Weppner, J., and Lukowicz, P. Activity recognition and nutrition monitoring in every day situations with a textile capacitive neckband. In *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication*, UbiComp '13 Adjunct, ACM (New York, NY, USA, 2013), 155–158.
12. Chowdhury, T. A., Purohit, P., and SA Fahim, M. Location based path guiding system for the visually impaired people. *Global Journal of Computer Science and Technology 14*, 1 (2014).
13. Hirsch, M., Cheng, J., Reiss, A., Sundholm, M., Lukowicz, P., and Amft, O. Hands-free gesture control with a capacitive textile neckband. In *Proceedings of the 2014 ACM International Symposium on Wearable Computers*, ISWC '14, ACM (New York, NY, USA, 2014), 55–58.
14. Ibañez-Guzman, J., Laugier, C., Yoder, J.-D., and Thrun, S. Autonomous driving: Context and state-of-the-art. In *Handbook of Intelligent Vehicles*. Springer, 2012, 1271–1310.
15. Levy, B. S., and Sidel, V. W. *Social injustice and public health*. Oxford University Press, 2013.
16. Sporka, A. J., Carson, B. L., Nauert, P., and Kurniawan, S. H. Toward accessible technology for music composers and producers with motor disabilities. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '13, ACM (New York, NY, USA, 2013), 72:1–72:2.
17. Sucu, B., and Folmer, E. Haptic interface for non-visual steering. In *Proceedings of the 2013 international conference on Intelligent user interfaces*, ACM (2013), 427–434.

18. Sucu, B., and Folmer, E. The blind driver challenge: steering using haptic cues. In *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility*, ACM (2014), 3–10.
19. Tomasi, C., and Manduchi, R. Bilateral filtering for gray and color images. In *Computer Vision, 1998. Sixth International Conference on*, IEEE (1998), 839–846.
20. Wang, F. T., and Hsiung, C.-C. A theorem on the tangram. *American Mathematical Monthly* (1942), 596–599.
21. Wobbrock, J. O., and Gajos, K. Z. A comparison of area pointing and goal crossing for people with and without motor impairments. In *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets ’07, ACM (New York, NY, USA, 2007), 3–10.
22. Xu, C., Israr, A., Poupyrev, I., Bau, O., and Harrison, C. Tactile display for the visually impaired using teslatouch. In *CHI’11 Extended Abstracts on Human Factors in Computing Systems*, ACM (2011), 317–322.
23. Zarin, R., and Fallman, D. Through the troll forest: exploring tabletop interaction design for children with special cognitive needs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2011), 3319–3322.
24. Zhao, H. Interactive sonification of abstract data-framework, design space, evaluation, and user tool.

Oh, That's Fancy! Methods and Techniques for Projection Mapping

Jonny Vang

ABSTRACT
Projection mapping, also known as Spatial Augmented Reality in research, is a technique which utilises one or more projectors to visualizes objects as projections on any devisable surface. Different from normal projection, projection mapping also works on non-flat surfaces, like the whole surface of a car. Because projection mapping includes the shape of the surface for its advantage, it is possible to create incredible illusions. For example with the help of this technique, it is possible to create astonishing set ups, like a gigantic Pac-Man game, which is projected on room walls, the ceiling and floor or the use of buildings like a family home as projection surface for a variety of projections to deceive the human perception. The advantage of projection mapping over traditional AR approaches is the non-use of devices between the user to display those computer generated graphics onto the real world. This paper gives a short insight on projection mapping. Therefore some issues will be discussed like a general description of projection mapping or its historical context. Furthermore different projects will be considered to understand, which possibilities exist to implement such projection mapping systems. Also the problem how to integrate those systems into the daily life will be broached.

Author Keywords
Projection mapping; spatial augemented reality; video mapping; augmented reality; depth cameras; immersion; illusion.

INTRODUCTION
Projection mapping becomes more and more popular and it is used in various ways nowadays. It is also implemented already in some real-life scenarios. Like using a variety of projections on concerts or music videos for flashier effects to attract the attention of the audience and deliver an astonishing experience for them. Or the use of projection on different shapes and kinds of surfaces like buildings or vehicles for promotional purposes. To generalize and expand those examples, projection mapping can be used in art, commercials or the entertainment section, but it is also an hot topic in research. They try to use projection mapping to enhance the overall user experience in daily life. Moreover it is an old dream of the human kind to merge the virtual and real world together to explore new ways of immersion, like it is shown in the holodeck of the Star Trek series. So it is not very surprising that people are working towards and improving such a technique like projection mapping. It became even more popular in the recent years [9]. For one thing thanks to the progress in the hardware development, namely the cheaper prices and consequently the affordability for such projects and for another thing thanks to the increasing interest of artists and advertisers for projection mapping around the world, because of the versatility of this technique.

Figure 1. Light Garden from teamLab - Astonishing example how projection mapping can be used in art. [23]

In the following this paper describes the motivation behind projection mapping and why it is used by such a wide variety of different people. Furthermore it explains the core elements of projection mapping and how it is actually related to Augmented Reality (AR). The historical context will be revised and also the different communities and branches in the industry which are using projection mapping will be discussed.

In particular this work explains various methods and techniques with help of different projects. For example what people are actually using to implement projection mapping set ups, plus still existing problems in the implemantation and the idea behind the project. Finally follows a discussion if those projection mapping installations could already be integrated into the daily life and the advantages and disadvantages of such a realisation. The last section also discuss further projects and ideas which could be implemented in the future.

Augmented Reality
Projection Mapping is actually an extension of Augemented Reality. The basic idea behind AR is to overlay real world objects with computer generated graphics. In other words, to integrate or rather merge the virtual world into the the real one. It will be implemented with the help of additonal devices

7th Seminar on Research Trends in Media Informatics (RTMI 15). February 2015, Ulm University, Ulm, Germany.

like head-mounted displays (HMD) or handheld devices with integrated camera. Those devices uses the live camera feed to display it with additional informations. An example could be to display a virtual person in the room. With the help of this approach additonaly information can be added to the real world scenario to enhance the user experience. Like displaying information in a museum as a visual guide. Also with the help of AR, illusions can be created to entertain the user. But those needed devices are quite an issue. First of all, all users, who should see the illusion, etc. need such a device. This situation is not everytime given, so only a handful of users can experience the AR simultaneously. Additonaly in case of the HMD the view of the user is rather limited. In case of an handheld device the device literally destroys the immersion of the user because the user has to hold the device between himself/herself and the real world. Thanks to those devices the interaction between the user and the displayed objects and informations is quite unnatural. Project mapping addresses those issues.

Projection Mapping

Projection mapping, more commonly known as Spatial Augemented Reality (SAR) or video mapping in research, is basically a technique which uses one or more projectors and any surface, be it flat or non-flat, as its canvas to show computer generated images. This is also the main difference between projection mapping and normal projection. Conventional a normal projection set up needs a flat surface to project onto it, otherwise the shown image or video will not be displayed correctly. Information could be lost or the projected object will be displayed in some distorted manner because the projection set up does not consider the surface and shape of the object. Whereas projection mapping takes full advantage of the geometric form of the surface itself to create astonishing illusions. Thanks to this concept we can enhanced the main idea of AR to get rid of those devices in between user and the real world. Therefore the interaction with the projected content becomes more natural and self-explanatory. Additionally projcetion mapping improves the feeling that the virtual reality is one with the real one.

MOTIVATION

Thanks to the dynamic and special features of projection mapping, it is used in many varying ways and in different industry branches and communities nowadays. Overall the different branches can be divided into arts, commercial, entertainment sections like movies or gaming and systems which improve the daily life and human-computer interaction in general.

Arts

One group of people which is very interested in such a technique like projection mapping and also responsible for the current boom of it, is the art community. With the help of projection mapping, artists can use the whole world as their canvas and create unprecedented art installations. Be it an huge building like a skyscraper, which displays a sort of an illusion or the human body itself to create similar or maybe better results and illusions like body painting. And it is not only the canvas aspect which is appealing to artists. Thanks to projection mapping artists can animate previous static objects and give them dynamic points. Or even further to play with the form of the surface itself to create unbelievable illusions and art installations. Also that projection mapping does not destory the canvas surface is a strong point. Therefore artists have access to surfaces they can work on, which were not accesible before because it would change the object permantly or even destroy it. Examples of such surfaces would be offical buildings or art objects itself. Furthermore the idea behind a projection mapping art exhibition is appealing too. Interestingly the art community and its projection mapping illusions influences the development of projection mapping heavily and respectively the other branches which are working with projection mapping. For example the project IllumiRoom [14], which will be discussed later on too, was influenced and inspired by art installations.

Commercial

Another branch which is working with projection mapping is the commercial one. The idea behind commercials is to attract attention of potential customers and create curiosity towards the specific product, so the customer buys the product eventually. Therefore projection mapping is quite fitting for commercials. For example they can project their product on a building surface or make a promotional video in which the product is transformed in an interestingly way with the help of projection mapping. This also applies to campaigns like welfare campaigns, which want the attention of people.

Entertainment

Besides those big installations, projection mapping can also be used in another way. The entertainment industry always looks to find new ways to improve the user experience and enjoyment. A concept could be to extend the static screen and transform the whole room into the scene. Therefore people would be immerse themself even more into the shown content. Another idea is to expand the shown content only with a few aspects, like a projected ranking list in a racing game. In addition the whole room could be the playground of a videogame with interactive elements to assure new user experiences and ideas for game developers.

General

Human and computer interaction also benefits from projection mapping. You could take use of the environement and interact with a system, which is more natural than a static display. Projection mapping could also be used while working. For example to display several screens [20] or it could be used to learn or support a manual work, for example to learn electronic modeling through projection mapping like described by Y. Akiyama et al. [2]. Now take that step even further and include a projection mapping system into an house. It could enhance the overall living standard. Wall decoration could be changed as desired or inforamtion could be displayed everywhere in the house to help the user in the daily life. Such displayed information would be for example a recipe which is projected onto the kitchen table while cooking.

HISTORY

It sounds like projection mapping is some recently discovered technique, but actually it started already 45 years ago. Disney was one of the first ones, which implement projection on a not-flat surface. They used 1969, in their Haunted Mansion Ride in Disneyland, 5 singing busts. To accomplish that, they filmed head-shots of the singers and projected them onto the busts [26].

The next known project was in 1980, Displacement by Michael Naimark [18]. This was an art installation, in which two people were filmed with the help of a rotating camera. The next step was to replace the camera with a projector which showed the taken video feed.

1991 Disney also made a patent of a system which used digitally painting onto "a countourd, three-dimensional object" [10].

An interesting date is 1998. Ramesh Raskar et al introduced the idea oh their Office of the Future [20] and shaped the academic term Spatial Augmented Reality. The idea behind this was to enhanced the work environemnt. Besides the computer monitor, we would also have augmented displays all over in the room with the help of projectors.

In 2001 Raskar et al. showed their Dynamic Shader Lamps [3]. This was a project in which the user can paint in real-time on real life objects with the help of projectors and a stylus.

Recent projects including Mano-o-Mano in 2014 [5], IllumiRoom in 2013 [14] or SurrondWeb, also in 2014 [24], are discussed later on in the paper. There are a lot of other not yet mentioned projects in regards to projection mapping, but mentioning them all would go beyond the scope of this paper.

METHODS AND TECHNIQUES SHOWN IN PROJECTS

Basically to implement a projection mapping set up, a projector and a computer which can link itself to the projecotr is needed. Actually every standard projector can be used for the set up. For a wearable set up pico projectors are used. The next step would be to know onto which surface the generated object will be projected to consider it shape and surface for the projection. It sounds quite simple at first, but there are different approaches how to map your generate images or animation to the real world. In the following various projects will be considered to understand their procedure and which methods and techniques they are using to understand the whole concept of projection mapping better.

Rendering Beforehand on a Known Canvas

The first and easiest method is to render the projection especially for an known object. Therefore the canvas object will be chosen beforehand. This means the properties shape and surface can already be considered, while creating the projection elements. Because the canvas is known, you just have to create objects which fit to the surface and map them digitally onto them. In the end the projector or projectors just display the prerendered graphics.

Found the House I grew Up in from FoundStudio (Figure 2) [22] is such a project. It's a commercial in which the team projected the story of the daily life of the residents onto the

Figure 2. The House I grew Up In from Found Studio: They take the form of the house into consideration and map their video feed beforehand [22]

house. After they found a appropriate house to project onto, they rebuild a model of that house in Cinema4D and used that model as their scene, on which they can add the animations and videoclips manually. They rendered each part specifically, with the position on the model in mind. After all clips were rendered, they are just adding all those onto the house model. The final product is a video which they just projected on the house in the end.

Another similar project is Box from Bot and Dolly [8]. They used a moving rectangular screen with the help of an robotic arm and projected their generated video onto the screen, while it was moving. The projeced video feed consist of various illusions. But this is actually not any different from Found the House I grew Up in. Because ,besides the canvas factor, they just need to know the movement of the surface beforehand, which is here the case. The whole set up is geared to each other.

A freeware Tool, which follows the same pattern and let someone project onto a static surface is VPT 7 [25]. It is like an image editing program and works with layers. It needs a projector and camera, both connected to the computer, to work. Then the camera films the wished scene or object and displays it for the following steps in the program. Now you can add manually layers onto the object, so that you can add as result the graphics and videos. It is a good entry point for someone who wants to start with projection mapping.

Capture Unknown Canvas

A more complicated approach is to create and render a projection onto an unknown surface. This may be the case, when the installation should be displayed correctly on different canvases. For example different rooms with different interior decoration. The problem is, the room structure and interior design is unknown, therefore it is impossible to remodel the rooms. As a consequence of this issue, the previous approach to map the generated objects manually is unsuitable for this case. Depth cameras like the Microsoft Kinect [17] in connection with the projector provide remedy. Depth cameras work with an Infra Red (IR) projector and an IR camera. The IR projector projects a pattern of IR light dots on all objects around it. Then the IR camera capture those IR dots. But this doesn't apply the depth sense, which is needed to remodel the scene. The actual IR video feed will send to an depth sensor processor. This processor works out the depth of the scene with the help of the placement of the projected dots. On near objects the pattern is spread out and on far ones it is dense. Thanks to that technique, it is possible to remodel the scene dynamically. Those projection mapping set ups just have to scan each new room or when the room structure is changed to work properly.

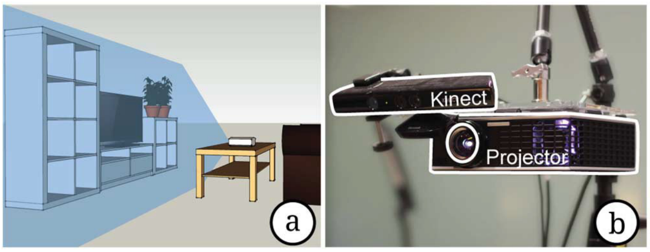


Figure 3. (a) Ideal setup of the IllumiRoom system. Using an ultra-wide field of view device that is placed on a table. (b) Their prototype setup, using a normal projector and Kinect sensor. [14]

Various projects are using depth cameras to capture the scene like IllumiRoom from Brett R. Jones et al.(Figure 2) [14]. IllumiRoom is a project, which displays in every conceivable room structure additional gaming content. For example to extend the view of the user or to illusions to improve the immersion of the user with the gaming content. Therefore the system needs an one time only calibration with help of the Kinect Sensor, like described before. It uses one projector and one depth camera as setup. The system itself has various projection illusions, which also are influenced by the ingame scenario implemented in the Unity3D engine. They come up with two concepts, how to integrate IllumiRoom into the game. The first one has direct access to the game itself through an implemented interface and extends the actual display. Thus the whole room itself is the screen and ingame content can be displayed in the room. For example the outline of the ingame scene can be extended and then displayed in the room. But to accomplish this the game has to be developed with the IllumiRoom system in mind and has to offer the interface. Because this is naturally not supported by every game, Brett R. Jones et al. come up with the another solution. They are using the controller input of the player to create different illusions, which are reacting to the input. They implemented illusions like falling snow or a moveable grid, which moves with the controller input or the whole room starts to wobble, when a certain button is triggerd.

Mapping on Unknown Canvas in Real-Time

Although the projection mapping installation has the scene captured thanks to the depth camera, it is still unknown where to project the objects, which are still rendered beforehand, in real-time on which surface. First of all with the help of the generated depth map, it exists a scene. Additionally with various algorithms like the Hough Transformation which finds planes in an environment, it is possible to distinguish dynamically between surfaces on which the objects can or rather should projected onto it. For example the floor would be logically the lowest plane in the scene. Each 3D point in the scene will then be associated to the nearest plane and can addressed later on from the system.

One project which has to work with this issue is RoomAlive from Brett Jones et al. [13]. It is a system in which the user can interact with various projections, which are projected onto a whole room surface. Test implications are games like fighting with robots which are projected somewhere in the room and can be controlled with the help of a seperate controller. Or playing the well-known game Whac-A-Mole, which project those moles somewhere in the room. The system set up consists of several depth cameras and projectors pairs positioned around the room, each connected to a computer. Because the mapping has to be in real-time and dynamically, they come up with four different solutions. The first and trivial idea is to project randomly. The system can map various objects on random 3D spots in the room scene. Another option would be to take the location into account. Thanks to the generated 3D scene in the system and the associated planes, grass, stones and other similiar object could only projected onto the lowest plane, which would be the floor. The other two solutions include the user himself/herself. One of them is to let the use tap a surface where the object has to be projected. The system recognizes the user's hand and projects the object to the 3D point which overlaps with the coordinates of the hand. The other idea takes the current location of the user into account and projects only onto the 3D points in front of him.

User Viewpoint

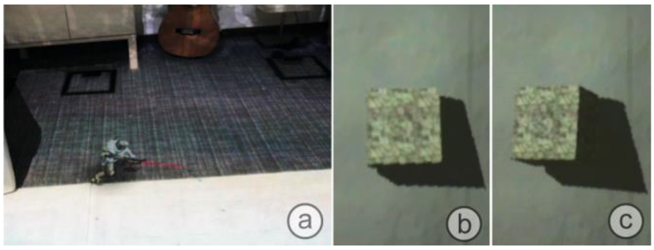


Figure 4. Different projections from different viewpoints from the RoomAlive project [14]

The next issues involves the user viewpoint. The user viewpoint is really important to maintain illusion of a projection mapping system. For example, the system project a cube on the ground and the user stands in front of the projected cube. From this viewpoint it seems like there is really a three-dimensional cube in the room. But when the user change his position and looks onto the projection of the cube from a different angle, the cube would look distorted. This is because the image of the cube is only generated to look only from this one viewpoint correct and the system still does not consider the viewpoint. Therefore the illusion would be destroyed. To prevent this to happen the system has to track from which angle the user looks onto the projected objects.

Brett Jones et al. had to solve this issue in RoomAlive too. Their system tracks the movement of the users head and implement those changes with the help of a two-pass view dependent rendering to display the object correctly from the viewpoint and changes it dynamically. (Figure 4). Virtual objects which belong to an physically object, like the wall surface change its appearance to a wooden one, have not to be rendered explicitly with the viewpoint of the user in consideration. The issue of more than one user in a room is not really solved in the RoomAlive system, because they just using the average head positions of all users, which only delivers good results, when object is near a physically surface.

User Tracking and Interaction

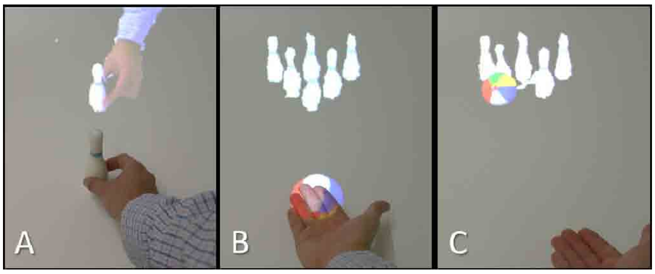


Figure 5. MirageTable - User helds a virtual ball in his hand and throws it against virtual pins. [4]

To interact with projected objects, like tapping, it is not enough to just track the user head movement. The system has to track the whole movement of the user. Also it has to exist some sort of collision interrogation. For example a ball is projected onto the users hand. To throw the ball and that the ball act accordingly afterwards in the room, the system needs to know how to react. Therefore the system needs a physic enginge. But a physic engine alone is still not enough for the system to recognize if the ball collide with the user hand or not. For this issue the system needs an internal model of the user, so that the engine can react accordingly to it.

MirageTable from H. Benko et al. [4] is one of those project which solved that issue (Figure 5). MirageTable tries to merge the real world with the virtual one by interaction with virtual objects and intgrate real objects into the virtual one. For example the user can scan a toyblock into the system with the help of the installed camera in the setup and copy it virtual into the system. Now the user can build a virtual tower with the blocks. The system now needs some sort of model for the users hand to recognize, that the user takes one of those virtual blocks. To solve this H. Benko et al. assimilate to the depth map of the hand proxy particles, tiny tightly packed spheres. Then those spheres participate in the physic simulation with other objects but not between the spheres themselves. They added also a force vector to each corresponding object, like the hand to improve the physic simulation. This whole idea also applies to the whole body of the user and not only the hand. Therefore proxy particles will simply be assimilated to the whole depthmap of the body of the user.

Multiple User

All those methods and techniques mentioned before, actually only work properly for one user in the room. This begs the question how to handle multiple users in a room. The Mano-o-Mano system from Hrvoje Benko et al. [5] addresses this issue. Although their system only considers only two users and not more. They used three projectors with each paired with a depthcamera in their setup. One projector faces user A, the second projector faces user B and the last one covers the room between them and faces the ground from above. All three projectors and specific depth cameras are calibrated, so that they create one scene and not three different ones in the system, which would lead to distorted and overlapping projections. The issue is solved with the help of set points in the room, which are shown in each video scene. The system now calibrates each video scene to the set points and merges it to one scene in the program. Now the system has one scene to work with, like in the previous projects.

Figure 6. To display projection to each user, the other user is used as a projection screen as well [5].

Another problem is to let both users interact with the same projected object. Therefore they have to be screens as well. For example user A holds a ball in his hand, therefore the ball is projected onto his hand. But so that user B can also see the ball in a properly way, the second projector has to project the ball simultaneously also on the body on user A. To extend the example, when user A throws a virtual ball to user B, it first has to be projected onto the hand of user A and then through the room to user B to simulate it in the correct way and maintain the illusion. The problem behind this project is, it only works, then both users face each other. If they dont, the projected objects can not displayed correctly.

IMPLEMENTATION TO THE DAILY LIFE

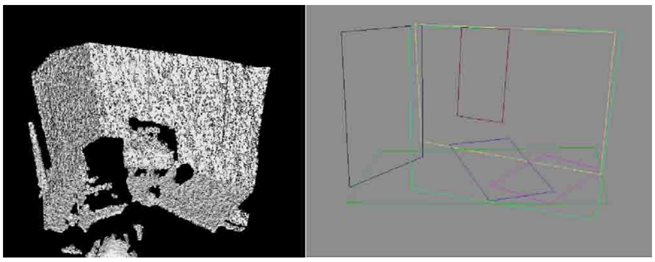


Figure 7. Use of the Room Skeleton model in SurroundWeb: On the left is the depthmap of the Room, on the right is the Room Skeleton [24]

Those systems could be easily integrated into the daily life nowadays. Getting the hardware itself is no problem because projectors and depth cameras in form of the Microsoft Kinect are quite affordable. Also like described in the different projects, it needs just an one time calibration for the system to work properly. Therefore such system could already in every hosehold. But there are still some problems. First of all projectors cannot achieve every desired color on every color surface. For example a projector cannot achieve a bright red color on a green surface. This is impossible at the moment. Another point is the fear of secruity loss, when the system needs to obtain information from the internet. Because most of those systems need depthcameras installed in the room. Therefore an outsider could hack those cameras to get access to the video feed. Similar to the uproar of the anoucement that each Xbox One is bundled with a Microsoft Kinect.

A current good solution offers the SurroundWeb system from John Vilk et al. [24]. This system is a browser, which utilise the whole room with the help of projectors and depth cameras as screen. With the help of the generated depth map of the room, they create a so-called Room-Skeleton Figure 7. This Room-Skeleton only provides information of flat surfaces saved as planes. Therefore the internet service only sees this model and not the actual video feed.

FURTHER PROJECTS AND IDEAS

There are still plenty ideas to use projecetion mapping in several way. Neclumi is a project from panGenerator [19] which creates projected jewellery with the help of a pico projector and a smartphone, which considers the gyroscope to change the jewllery dynamically. The next step from projection mapping would be probably to create real 3D hologramms in a room. There is actually a project called Bleen [7], which promised hologramms from an obscure device. The problem is they want to project the 3D model into the air and this is not possible with the current technology, so it is unfortunaly a fake.

Other ideas for further projects could be to develop a new communication device with the help of projections. The current devices are not really optimized for human conversation, when you take into account that we communicate through displays. It would be more natural, when the conversational partner is projected onto a surface, so that is seems he is actually in the same room.

There is also the possibilty to use projection mapping to shape the interior design, without damaging the walls or furniture. Instead of buying a picture and a frame for it and hanging it onto the wall, you could just project the image. Another positive aspect is that you can change the picture every day. Or imagine you could have your own real time waterfall on your wall. Considering that, interior designers could design the room with customer at the same time in the room with immediate results how any design proposals appears.

Another scenario could be while shopping clothes, you could at first project the clothes onto your body to see if it actually matches with your style before try the clothes on.

Also think about projected guides in big buildings or museums with the help of mounted projectors all over the place in the building. Thanks to that people would not get lost at new places.

A more useful idea would be to use it as a kind of tutorial for new user. For example for car repair. You could project each step onto the real object itself, like to screw at this specific part or a projected arrow shows where to change the oil for example, and achieve the whole repair by your own, without the help of a mechanic. This of course is not limited on car repairs.

SUMMARY AND CONCLUSION

Projection mapping is some astonishing technique which enhanced the user experience in several ways. Like the paper discussed it can be used in arts, just for entertainemnt purpose or for more constructive concepts, like displayingionaditonal informations. It is not limited to the surface at all like normal projection and therefore not that limited in execution. The future will shown how projection mapping will be further integrated into the daily life and how the technique further improves. Maybe it is a beginning to understand how to create 3D holograms in the future.

REFERENCES

1. Adcock, M., Thomas, B., and Feng, D. Visualization of Off-Surface 3D Viewpoint.
2. Akiyama, Y., and Miyashita, H. Projectron Mapping : The Exercise and Extension of Augmented Workspaces for Learning Electronic Modeling through Projection Mapping.
3. Bandyopadhyay, D., Raskar, R., and Fuchs, H. Dynamic shader lamps : painting on movable objects. *Proceedings IEEE and ACM International Symposium on Augmented Reality*, 207–216.
4. Benko, H., Jota, R., and Wilson, A. MirageTable: freehand interaction on a projected augmented reality tabletop. *Proceedings of the SIGCHI conference on ...* (2012).
5. Benko, H., Wilson, A. D., and Zannier, F. Dyadic projected spatial augmented reality. *Proceedings of the 27th annual ACM symposium on User interface software and technology - UIST '14* (2014).
6. Bimber, O., and Raskar, R. *Spatial augmented reality: merging real and virtual worlds. AK Peters*. 2005.
7. Bleen. Bleen, Janurary 2015. https://bleen.com/de-de/.
8. Bot&Dolly. Box, Janurary 2015. http://vimeo.com/75260457.
9. Central, P. M. Projection mapping central, Janurary 2015. http://projection-mapping.org/.
10. Google. Apparatus and method for projection upon a three-dimensional object, Janurary 2015. http://www.google.com/patents/US5325473.
11. Hardy, J., and Alexander, J. Toolkit support for interactive projected displays. *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia - MUM '12* (2012).
12. Irlitti, A., and Itzstein, S. V. Validating constraint driven design techniques in spatial augmented reality. *Proceedings of the Fourteenth Australasian User ...*, Auic (2013).
13. Jones, B., Sodhi, R., Murdock, M., Mehra, R., Benko, H., Wilson, A. D., Ofek, E., Macintyre, B., Raghuvanshi, N., Shapira, L., and Tech, G. RoomAlive : Magical Experiences Enabled by Scalable , Adaptive Projector-Camera Units.

14. Jones, B. R., Benko, H., Ofek, E., and Wilson, A. D. IllumiRoom : Peripheral Projected Illusions for Interactive Experiences.
15. Maas, E., and Marner, M. Supporting freeform modelling in spatial augmented reality environments with a new deformable material. *Proceedings of the ...* (2012).
16. Marner, M., and Thomas, B. Spatial Augmented Reality User Interface Techniques for Room Size Modelling Tasks.
17. Microsoft. Microsoft kinect, Janurary 2015. http://msdn.microsoft.com/en-us/library/hh855355.aspx.
18. Naimark, M. Displacements, Janurary 2015. http://www.naimark.net/projects/displacements.html.
19. Neclumi. Neclumi, Janurary 2015. http://www.neclumi.com/.
20. Raskar, R., Welch, G., Cutts, M., Lake, A., Stesin, L., and Fuchs, H. The Office of the Future : A Unified Approach to Image-Based Modeling and Spatially Immersive Displays. 1–10.
21. Ridel, B., Reuter, P., and Laviole, J. The Revealing Flashlight: Interactive spatial augmented reality for detail exploration of cultural heritage artifacts. *Journal on Computing ...* (2014).
22. Studio, F. The house i grew up in, Janurary 2015. http://www.found-studio.com/case_study/hiscox/.
23. teamLab. Light garden, Janurary 2015. http://www.team-lab.net/.
24. Vilk, J., Molnar, D., Ofek, E., and Rossbach, C. Least Privilege Rendering in a 3D Web Browser.
25. VPT7. Vpt7 - freeware, Janurary 2015. https://hcgilje.wordpress.com/vpt/.
26. Wikipedia. Grim grinning ghosts, Janurary 2015. http://en.wikipedia.org/wiki/Grim_Grinning_Ghosts.
27. Wilson, A., Benko, H., Izadi, S., and Hilliges, O. Steerable augmented reality with the beamatron. *Proceedings of the 25th annual ACM symposium on User interface software and technology - UIST '12* (2012).
28. Wilson, A. D., and Benko, H. Combining multiple depth cameras and projectors for interactions on, above and between surfaces. *Proceedings of the 23nd annual ACM symposium on User interface software and technology - UIST '10* (2010), 273.
29. Zhou, J., Lee, I., and Thomas, B. Applying spatial augmented reality to facilitate in-situ support for automotive spot welding inspection. *Proceedings of the 10th ...* (2011).

Microinteractions for Wearable Computing

Jörn Hofschlaeger
Institute of Media Informatics
Ulm University
Ulm, Germany
joern.hofschlaeger@uni-ulm.de

ABSTRACT
In this paper I want to explain what microinteractions are and why we need them. There are multiple ways to use wearable devices, but still problems with input possibilities to solve. The idea of microinteractions is to minimize the time of usage without changing what the users reach or want. Therefore I present different input methods which can be used to reach the wanted targets faster and explain what can be used for which scenario and why it cannot be used for the others. I conclude with future possibilities and how the new potential ways of interacting can help to use some of the currently not in everyday life usable devices.

Author Keywords
Microinteractinos; mobile devices; interaction techniques; gestures.

INTRODUCTION
In this paper I want to explain some things about microinteractions and why we need them. A microinteraction is every interaction which takes less than four seconds to initiate and complete where it is not relevant which device one is interacting with.

The reason to work with microinteractions, when working with mobile devices is that most people can maintain concentration to everything around them and the mobile device just for bursts of four to eight seconds [18]. So everything longer than a microinteraction could distract someone from things around them. So the main task of microinteractions is minimizing the time the tasks take, so one can go back on the task at hand [2]. Furthermore according to Cui *et al.* many people miss their phone calls while carrying their mobile phone with them just because of the way they carry it [6]. To sum it up, we need new devices or something new supporting the existing ones.

Ashbrook differentiate in his dissertation [2] between two phases of a microinteraction: access time and usage time. While access time describes the time to get the device in the right place to use it from its storage location and direct to the wanted application, usage time is the time which is needed to get the application to do what you want, which can take everything from just a few seconds to several minutes. So he decides to just consider access time for deciding if it is a microinteraction or not.

We need more ways to create microinteractions, because we use more and more devices and they distract from what we are actually doing. Especially because there is a big lack of immediacy. Even when we carry a smartphone, it takes a while to take it out of the pocket and to unlock it. According to Starner it takes about 20 seconds from the intention to use the smartphone until it is ready to be used [21]. This means that even a trivial thing like reading an incoming Message, what just is a microinteraction by the definition of Ashbrook, takes these 20 seconds. So it is very important to not forget thinking about the interfaces [16] because a desktop desktop-computer user normally uses a mouse and a smartphone user does not, because for him it is much easier to use the touchscreen. So it allows different interactions and creates also not the same problems e.g. it is nearly impossible to tab with the finger on a very little icon but for that it is easier to get from one end to another to click there something. This new interfaces are today implemented nearly everywhere and allow new ways of interacting with them.

Another and even more current problem in research is, that users of mobile devices do not want to keep standing or sitting while using their device and it is much harder to click on a small button on a touch-screen when one is moving. Though this way of interacting is not the best for every situation [19]. Another technique might be better in this case, but even eye free devices lose a lot of accuracy when the user is running while using it [11]. Ashbrook discovered something which is on the first view very contrary: He tested the device access time of going and standing people and found out that it makes nearly no difference i.e. in his test the going people have been a bit faster than the other participants who were standing [3], comparing this results with what Schildbach and Rukzio revealed in [19] shows that there is a big difference between access time and the higher usage time which is owing to the additional time is needed to tab on the exact right space.

Another paper [17] shows the impact of carrying something in hand i.e. a shopping bag and meanwhile using a smartphone. Ng, Brewster and Williamson tested this scenario and let the subjects using a smartphone with one- or both hands i.e. in landscape and portrait orientation. They found out that whatever method one is using it is harder to use the smartphone encumbered i.e. the main target accuracy is at least 10 per cent lower.

In the UK there have been 6.6 million accidents in 2007 [7]

7th Seminar on Research Trends in Media Informatics (RTMI 15).
February 2015, Ulm University, Ulm, Germany.

Figure 1. Nenya [1]: A technique using the magnetism of a ring to produce input for a wearable device. For this the receiver on the arm can interpret the changes in the magnetic field.

which can led back to smartphone usage and some of this accidents have been broken noses and even a fractured skull. To prevent by wearable devices distracted people to injure themselves Britain got the idea of padding lampposts, so even if they run into them, they do not get injured [7]. So I want to present some new techniques using different ways to make interaction with wearable devices easier and faster so the distraction is shorter and hopefully not longer than four seconds.

DIFFERENT WAYS TO CREATE MICROINTERACTIONS

There are multiple possibilities to create microinteractions with different possible use cases and problems. I categorize these interactions with the underlying technology and discuss possibilities and limitations.

Magnetics

One way to create microinteractions is via magnetics like Ashbrook *et al.* did for their project Nenya [1]. For their research they took a ring which one can wear on finger and rotate and slide it to change the magnetic field and thereby generating the input for the device e.g. a smart-phone or -glasses (Figure 1). The receiver at the arm is able to get and interpret the differences in the magnetic field of a normal but magnetic ring when turning it. So no special ring is needed. However it is easier to use a ring with a landmark like a stone because so it is easier to go back on beginning. It is furthermore easier to know where in the menu one is at the current moment. This way of interaction is very discrete so one could change the song one is actually listening or the thing one can see on the smart glasses in a crowded place like in subway, without someone around seeing it. So this is already a reason to use this technology, another cause is that it is a way to interact eye free and the device itself is unpowered.

Harrison and Hudson tried something different with magnetics and called it Abracadabra [10]. They invented a way to extend the input area of devices having a very small or even no screen, so it is much easier to interact with them. For this they placed a small magnet at a finger and a magnetometer as receiver in an wristwatch. With this two devices it is possible to track the finger position everywhere around the device, so it is possible to control the smartwatch without touching it. Alike Nenya the magnets are unpowered. "Not only does this mean they never need to be recharged, but also enables them to be small and robust against impact and liquid damage" [10]. Unlike Nenya this way of interacting is not so discrete but for that there are more ways to interact than there are with [1] with which one normally just go through a menu or list.

A problem with Nenya and Abracadabra is that they cannot work alone because they need a device which receives the movement of the ring or the magnets on the finger. This makes it necessary to wear another device which looks like a bracelet with sensors or a smartwatch on it, which could change in future versions so one would just need the bracelet. Furthermore they had the problem that their bracelet or watch looked a bit odd because it needed to be bigger than it would have usually been since the additional sensors were not very small-sized. If one want to implement a magnetic sensor today it would not be such a problem because the current sensors are much smaller.

Cameras

A different way to create microinteractions is with cameras. There are multiple research teams trying to make interacting with devices easier, one of them is Kangas *et al.* [13]. This project is about gaze gestures for mobile devices combined with vibrotactile feedback. The idea is, that it is much easier to complete a task with gaze gestures when you get any kind of feedback. To make this possible Kangas *et al.* used a Tobii T60 eye tracker (Figure 2) and a Nokia Lumia 900. The testers had to navigate only with gaze gestures through a contact list on the phone and were getting vibrotactile feedback. To find out when people expect the feedback and so helps to user the most noticing that the gaze tracker noticed the gesture, they tried four different conditions for when the feedback gets triggered:

- No haptic feedback.
- Haptic feedback given when a stroke, originating from inside the device to outside the device, was recognized.
- Haptic feedback given when the second stroke, originating from outside the device to inside the device, was recognized.
- Haptic feedback combining the last two.

It is shown that the probands had the best results when they got the second way of feedback and similar results with the last test-configuration. Another thing Kangas *et al.* have evaluated is how many unnecessary gaze gestures have been made. Here it can be seen that every feedback lowers the rate of these unnecessary gestures at least by 14%. Feedback in every form helps increase efficiency and the experience for the testers, anyhow gaze gestures are still not easy to use. For

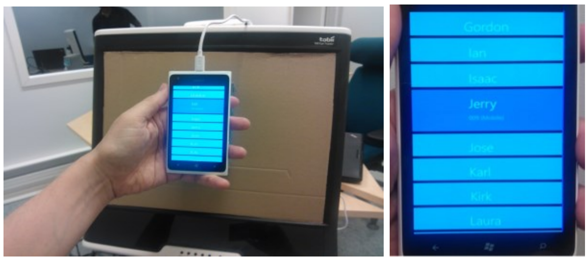


Figure 2. [13] A participant uses the Nokia in front of the Tobii eye tracker (right). The list he has to navigate through with gaze gestures (left).

controlling any kind of device, gaze gestures can only be used for very easy tasks like scrolling a screen which can thereby be done hands-free.
Another way to create microinteractions with camera is Weigel, Mehta and Steimle's [22]. They had the idea of getting input from finger gestures on the upper limb and for this they took some cameras which helps a microcomputer to understand multiple gestures. They chose the arm and hand as "input space" because it is highly social accepted [22]. Because there are multiple very natural gestures like for example touching, pulling and squeezing, it is very easy to control a mobile device and reach every wanted target fast and easily. During their studies they discovered that users often transfer already known multi touch gestures to on-skin input. This made Weigel, Mehta and Steimle using this known gestures as a already established standard. A problem is the placement of the cameras and maybe clothing because in the tested setup it works only with bear arms, but because it was summer when Weigel *et al.* tested it everyone wore short sleeve clothing or could easily uncover their arms. Another limitation is that participants were seated in the user study. In a standing or walking condition, gesture detection might be more difficult.

Depth cameras

Depth cameras allow other ways to create microinteractions. Gustafson *et al.* used this technology for their paper [9]. Here they wanted to control an Iphone without using the touch-sensitive display or any buttons, just by doing the gestures in the palm of the own hands. For this the depth camera tracks the movements and sends them wirelessly to the actual phone which interprets them like they would have been done on the phone itself. The idea is clearly not to control a smartphone, even when they did this for testing purposes. This technology is made for using it with screen-less devices but it was easier to test everything with commonly well-known devices. So while testing the participants had to try to do a every day task, like setting an alarm clock, without seeing the phone and keeping it in their pocket. To test their idea Gustafson *et al.* let the testers first do transfer leaning. For this they

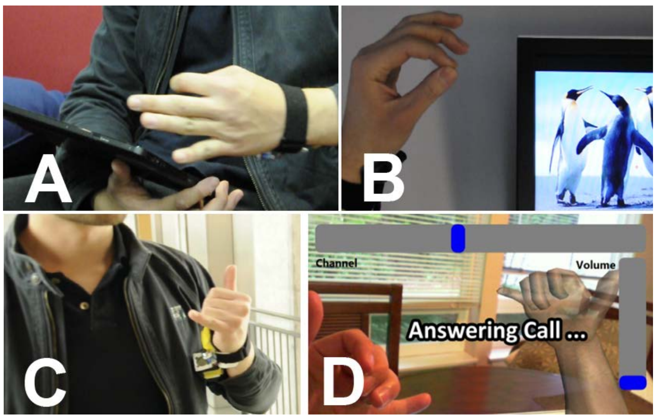


Figure 3. Digits [15]: Digits application scenarios. A+B) Extending interaction space around a mobile device into 3D. C+D) Nonvisual UIs allow users to manipulate application parameters without looking at or touching a physical device (GUI elements are for illustration only).

Figure 4. ShoeSense [4]: A shoe with an attached depth camera.

gave their probands a paper phone, so it is easier for them to get along with not touching on the real phone. Later they had to try it without the paper phones directly on the hand which is not this easy without seeing what happens but the idea of the authors is, that one does not need to see because one does already know.
Kim *et al.* tried something else with depth cameras they called Digits [15]. Their idea was to interact with the phone using just one hand and 3D interactions. For this, they wore wrist worn depth cameras, so they could track the hand movements to interact with a mobile device. The problems they had to face are very complex, so they had to solve the problem of differencing the fingers and ignoring the background. For this they found a way, with new algorithms and biomechanical knowledge of the hand, to build a complete 3D model of the users hand. Additionally they had to face problems like computational costs, form factor and power consumption.
Because Digits is one handed and very easy to use one can utilize it for a lot of applications including eye free interaction and while moving. A problem they could not solve is that depth cameras are neither small nor really unobtrusive. In contrast they solved the problem of energy consumption by using a depth camera with lower resolution.
Bailly *et al.* had the idea of using a shoe mounted camera which is up-facing to see hand gestures and called it ShoeSense. After using a Kinect for the first prototypes they looked for something else because this camera is very much too big for wearing it on a shoe. They looked for another depth camera which is smaller and needs less energy (Figure 4) [4]. The idea is, that the user has not to take his phone to hand but can use his phone without this by doing gestures with his hand. Where the idea is to use it while going at any given situation. Here they implemented three gesture sets: Triangle, Radial, and Finger-Count which can be used without any visual attention. However they had some problems like the size of the battery and the energy consumption of the depth camera which they could only partly solve.

Sensors

Xiao *et al.* used Sensors for their paper [23] in which they wrote about a method to expend the expressivity of smartwatches. According to them there is a need for additional input methods for smartwatches because of the very small display size and so it is very hard to use its display like a smart-

phones for every interaction. So they thought about panning, twisting, tilting or clicking the smartwatch itself. This allows very understandable interaction for multiple applications e.g. Xiao *et al.* used the clicking for switching from looking on the clock to the alarm clock and the twisting (clock and counterclockwise) to adjust the alarm time. In another application they used the panning to move on a map, the twisting to zoom in and out and the clicking to show additional data. A big problem of this implementation in a smartwatch is that very big components are needed to ensure the new functionalities which additionally make the device vulnerable to weather and sand. The idea of Xiao *et al.* to solve this big issue is to make the interaction smaller i.e. use sensors that get which movement would be made if it was possible. When doing this it would be not so easy to understand the new possibilities like panning and twisting.
Another method to create microinteractions is tested by Karrer *et al.* called Pinstripe [14]. They implemented sensors in clothes, so one can eye free control ones mobile devices very naturally and unobtrusive. In general the idea of smart clothing is not new but often implemented using buttons or textile UI elements that are hard to be used eyes-free. In Pinstripe, Karrer *et al.* had the idea of not implementing buttons or textile UI elements like that because it is, according to them, very hard to use them eye free. Pinstripe uses the fact that most closes lose folds when worn, so pinstripe can track folds and their approximate size everywhere the sensors are implemented in the clothes which could be everywhere and makes it possible to use this folds to navigate through a menu or use it as a button. The idea of just using folds for the input makes it much harder to interact accidentally with Pinstripe and because the sensors can be placed on the inside of the clothes one can implement them without having to make them noticeable. A problem with this idea is that it supports not much different possible input, so it is not usable for everything but one can navigate through a list or take a call.
A Further paper, written by Gong *et al.*, is about PrintSense [8]. This technique allows to print flexible sensors very cheap and easy. This printed surface is very easy to connect to any wanted hardware and allows to add an capacitive touch module to everything. Furthermore this printed sensors are able to act when they are folded, i.e. when added to a book or a newspaper it would be possible to highlight the next page. Even being an capacitive touch module it is possible to differentiate between three different levels of pressure [8]. However it is not possible to use this technique for much more than big flexible buttons. So this sensors are usable in just a few situations but because of the low price they are good for any kind of prototyping.

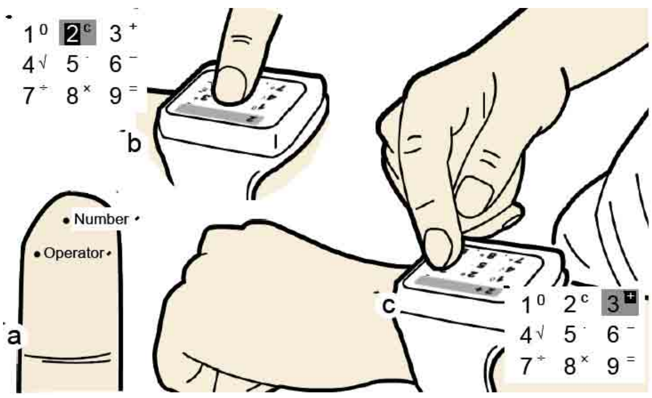


Figure 5. Touchsense [12]: a) The user can switch between typing operators and numbers by using different areas of finger pads. b) Using a normal tab '2' is selected. c) Using the right side of the finger '+' is entered. The gray background shows which field is touched and the black field shows what is entered.

Figure 6. Skinput [11]: The sensing armband augmented with a picoprojector which allows to see the interactive menu on skin.

Huang *et al.* tried something different in [12]. They had the idea of using varied finger postures to activate different functions on a smartwatch because this would minor the problem of having just a very small input area using the screen. To make this possible the researchers put a sensor on the finger and an additional one on the smartwatch, where the idea is that in the future the system can identify the postures with an in the touchscreen implemented fingerprint scanner. In their paper Huang *et al.* tested their system using a calculator application where on the front of the finger one use the numbers and with the side the operators (Figure 5). So with this technique one could enable an additional input layer on every current device, were it is more usable on devices with small screens because it here it would be cheaper and is more needed.
To sum it up sensors open very broad field and allow many different possible techniques to create microinteractions. Furthermore, they are very easy utilizable even in current state of the art. Additionally they are small and need little power which makes them easy implementable in many other devices.

Sound

The last technology, I want to present, that can be used to create microinteractions is by sound. The idea of Harrison *et al.* for Skinput [11] was to use the skin as input surface because it is very large and mostly easily accessible. For this they invented a device which listens to the sounds created by tapping on the arm. Where they do not use the sounds going through the air but them going as vibrations through the arm itself. Doing this Skinput can distinguish between up to ten tabs on different places on the forearm and hand with a high accuracy. Also it is possible to interact only with finger tabs on the wanted places and because the sound is different when tapping against the arm or hand with something else than a finger, it is possible to only do something when it is wanted i.e. the tab it really made with a hand. Everything harder than a hand e.g. wood or something soft like an LCD produces another sound an this makes it possible to use a system like this using just a single hand. These gestures are easier doable eye free but cannot be so soft like for a camera system because this would not create a loud enough noise.

Due to the variety of input it is possible to play games with [11] or just to control the music while jogging or walking and doing this eyes-free and discrete. Additionally one could combine this technique with a projector so it is possible to see the menu while using it directly on skin, which makes it a lot easier to tab on the right space, that makes it of course impossible to use while moving except the projector would move together with one i.e. is also attached to the arm what makes the device even bigger (Figure 6).

In addition Skinput is not really usable in every day situations, except maybe in summer, because it can just be used having a bare arm. Otherwise the created sound would not be like it is expected and so cannot be used like it is wanted.

GENERAL PROBLEMS IN CURRENT RESEARCH

A big problem in current research is size. To use most of the presented techniques one needs cameras [9, 13, 15, 22] or in case of magnetics something that gets the changes in the magnetic field [1, 10]. . The required sensing technology is often so big, that it is difficult to hide. To solve this we need time to invent technologies which are smaller and as good as the actual ones. Another way would be to work with sensors instead of cameras because they can be much smaller e.g. [8]. Making the new techniques invisible for others is right now more easy than getting everyone to like it and also it needs less attention from the user itself.

Another important value is acceptance and ethical concern which are related to the first problem: size. Here is the task to get people to like things they mostly do not like because they feel intimidated or injured in their privacy. So for example multiple public spaces in the US do not allow Google Glass and other smart glasses because the people could always film, photograph or record everyone around them without one could see a difference or even when they are not sure if Google could always films them. So someone wrote a script to block Google Glass users and everything what could be any kind of smart and so track things from Wi-Fi in public spaces or even at home. Some people founded an initiative called "Stop the cyborgs"[1] against all smart devices to protect themselves against Big Data and save their privacy. For this they created a sign to let everyone see that no smart glasses are tolerated inside which shows a crossed smartwatch. Also there are some people who are so feared by smart devices that they attacked a Canadian professor who wears permanently attached smart glasses and did not want to take them of when he got asked to because he cannot. When the attacker noticed that it will not come off this easy he tried it harder instead of letting the professor in peace [5]. This could get a bigger problem for even more people when it is not possible to see when someone interacts with his or her smartphone, smartwatch or glasses. To sum it up it is not easy to say what is better, on the one hand when making the possibilities to interact with mobile devices less noticeable someone will have a problem with it on the other hand many people do not like to put the smartphone out of the pocket, just to switch to the next song.

Furthermore there is the issue of energy consumption which is highly connected with efficiency. Here we have the problem on nearly every current device e.g. the battery of a smartphone just suffices for a day and the same problem we have on other wearable devices like smartwatches, except pebble[2] which uses an e-ink display. This is for some people a big reason not to buy smart wearable devices, because they would have to charge a device very often.

In addition to the already mentioned problems there is the big issue of distraction. Many current mobile devices use visual feedback and can hardly work without it. Starner writes in [20] about the problem of using wearable computers while driving or walking and imparts the idea that " a wearable interface should be designed to maximize performance and minimize investment inattention" [20]. The same problem Starner had in 2002 with head up displays like navigating systems in a car or smartphones while driving. Of course we use navigating systems in nearly every car, but it is not allowed anymore (at least in Germany) to change something with the hand on it while driving because it is considered to cause too much distraction for the driver. So we cannot use every technology at all time or at least are not allowed to. Some are better to control complex things but a task like switching to the next song while driving should be solvable eyes-free and not need seconds for the search of the right button.

Moreover there is the problem of finding an alternative to touchscreens which take less concentration and also being able to interpret so many different things. This is needed very often e.g. for textual input. The most alternative input mechanisms support just up to ten different inputs e.g. [11] and most even less like [1,10,15,23]. The only current alternative input device to the touchscreen is the voice e.g. Siri on Iphone, Cortana on Windows phone and Google voice on Android which are getting better and better but still cannot understand everyone in every situation e.g. in a noisy surrounding it is difficult to differentiate voice from the background. Additionally they have the problem that one have to speak to them very loud and it is not always wanted, at least by the most people, that everyone around can hear what one is "writing". Moreover it is just a good input for already known words. So if someone uses slang words or want to write a password which is not a word or sentence, even if it would be a sentence.

FUTURE WORK

Like I pointed out in the last section there are multiple problems in current research. Some of them will go missing like

[1] http://stopthecyborgs.org/

[2] https://getpebble.com/

the size others like energy consumption and efficiency not. There are always new problems or the ones we think we have solved change a little and we need new ideas again.
The problem of energy consumption is, that every time we have a solution for this problem i.e. making something in the devices more efficient, someone thinks about new gadgets he could implant and use with the now remaining energy so the problem cannot be solved like this. So we need better batteries but there is always something one can do with the energy which seems better than letting the users have more time before they have to recharge. However this problem cannot be solved like this because many people prefer new features over a longer battery life. This is of course not just valid for smartphones but also for every device and so also for other wearable devices.
The problem of size is very different. For example mobile phones at first were getting smaller and now they get bigger again while becoming thinner. This is of course because current mobile phones have touch-screens and it is easier to touch something on a big screen than on a small where the earlier mobile phones have had buttons and so it has been a bigger problem to get the phone in the pocket than hitting the right button. This means, that it is not always easy to decide if even a mobile device has to be smaller than it is currently because if we change the way to interact with them the wanted values shift. In other cases like the bracelet of [1] the size has clearly to decrease before it can be used in everyday life.
Another thing will change in future work is that projects heading in the same direction as [12] will get the possibility of using fingerprint scanners which are implemented in the touchscreen to make interaction faster and easier so one will not have to hit a very small button and for that using different areas of their fingers. This will make it much easier to navigate and generally using a smart device with a small input screen but touchscreens have the problem that it is hard to use them while moving and even when there are no studies to this until the technique is possible, it will be very hard to use them while moving and nearly impossible while jogging. This is of course because, like I pointed out earlier, it is already hard to use a current smartphone while moving [19] and the displays of smartwatches are much smaller than this.
Furthermore one could combine some things discovered by different projects. It could be an improvement to use the haptic feedback of [13] with [9] so it is easier for the users to sense that they are doing the thing they want. Another idea would be to add haptic feedback to [1] e.g. when the last item in a list is reached. Another interesting thing would be to add a display to [9] phone so it is possible to use this technique also for not well known things. This display would of course destroy a part of the concept but if it is just a display e.g. glasses or maybe contact lenses it would add many possibilities to the actual idea.
Additionally one could take the next step on [10] and implement the magnets so they are invisible and permanent available. There are multiple things one can do with magnets e.g. telling if something is magnetic or not and feeling magnetic fields and also to tell if a wire is live or dead. So it is mostly not something one really needs but it could make things easier especially if it is used for input on mobile devices too. Implementing things in the body one can get problems, in case of implementing something visible like the professor I mentioned earlier people felt intimidated.
Moreover in future work the gaze trackers will get better which makes it possible to use gaze gestures like in [13] in phones. There are already smartphones, e.g. Samsung S4, having a head-tracker which is a precursor of a gaze tracker. This shows that this feature is wanted, but the current precursor does not use haptic feedback like it is presented by Kangas *et al.* [13]. In current devices the head-tracker is used for scrolling. Furthermore it is used for keeping the display active while the user is looking on it. However it does not work very often and so is not used by the most people who have a phone with these technology inside.

CONCLUSION

In this paper I presented multiple new ways to create interactions and some of them are really good to establish microinteractions and make interacting in general and especially while moving easier and also a lot faster. However there are still multiple problems to face before most of this techniques can be used in every day situations. Some of the current projects are currently to big to use them outside of a lab environment (e.g. [9, 13, 15]) and some others are not implementable in smart devices like the authors planned (e.g. [23]) because the new sensors would make the device vulnerable against external influences. Although there are some projects, mostly with sensors, which can be used in everyday situation. Even in the current state of the art would be utilizable in every day situations e.g. [8, 11, 14] but they all have in common that they can just be used for few different inputs. So the most of them can just be used for navigating through a list or in case of [8] to switch something on and off and maybe switching between different modes.
Nevertheless we are on a very good way to make interacting with wearable computing devices easier and faster. Additionally it becomes a trend to really look if a device for interacting is usable while walking especially when they are made to use with wearable devices.

REFERENCES

1. Ashbrook, D., Baudisch, P., and White, S. Nenya: Subtle and eyes-free mobile input with a magnetically-tracked finger ring. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, ACM (New York, NY, USA, 2011), 2043–2046.

2. Ashbrook, D. L. *Enabling mobile microinteractions*. dissertation, Georgia Institute of Technology, 2012.

3. Ashbrook, D. L., Clawson, J. R., Lyons, K., Starner, T. E., and Patel, N. Quickdraw: The impact of mobility and on-body placement on device access time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, ACM (New York, NY, USA, 2008), 219–222.

4. Bailly, G., Müller, J., Rohs, M., Wigdor, D., and Kratz, S. Shoesense: A new perspective on gestural interaction and wearable applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing*

Systems, CHI '12, ACM (New York, NY, USA, 2012), 1239–1248.

5. Bilal, W. Cyborg america: inside the strange new world of basement body hackers. http://www.dailymail.co.uk/news/article-525785, Aug 2012. (Accessed: 17.12.2014).
6. Cui, Y., Chipchase, J., and Ichikawa, F. A cross culture study on phone carrying and physical personalization. In *Proceedings of the 2Nd International Conference on Usability and Internationalization*, UI-HCII'07, Springer-Verlag (Berlin, Heidelberg, 2007), 483–492.
7. DailyMail. Brick lane made britain's first 'safe text' street with padded lampposts to prevent mobile phone injuries. http://www.dailymail.co.uk/news/article-525785, Mar 2008. (Accessed: 17.12.2014).
8. Gong, N.-W., Steimle, J., Olberding, S., Hodges, S., Gillian, N. E., Kawahara, Y., and Paradiso, J. A. Printsense: A versatile sensing technique to support multimodal flexible surface interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, ACM (New York, NY, USA, 2014), 1407–1410.
9. Gustafson, S., Holz, C., and Baudisch, P. Imaginary phone: Learning imaginary interfaces by transferring spatial memory from a familiar device. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, UIST '11, ACM (New York, NY, USA, 2011), 283–292.
10. Harrison, C., and Hudson, S. E. Abracadabra: Wireless, high-precision, and unpowered finger input for very small mobile devices. In *Proceedings of the 22Nd Annual ACM Symposium on User Interface Software and Technology*, UIST '09, ACM (New York, NY, USA, 2009), 121–124.
11. Harrison, C., Tan, D., and Morris, D. Skinput: Appropriating the body as an input surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, ACM (New York, NY, USA, 2010), 453–462.
12. Huang, D.-Y., Tsai, M.-C., Tung, Y.-C., Tsai, M.-L., Yeh, Y.-T., Chan, L., Hung, Y.-P., and Chen, M. Y. Touchsense: Expanding touchscreen input vocabulary using different areas of users' finger pads. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, ACM (New York, NY, USA, 2014), 189–192.
13. Kangas, J., Akkil, D., Rantala, J., Isokoski, P., Majaranta, P., and Raisamo, R. Gaze gestures and haptic feedback in mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, ACM (New York, NY, USA, 2014), 435–438.
14. Karrer, T., Wittenhagen, M., Lichtschlag, L., Heller, F., and Borchers, J. Pinstripe: Eyes-free continuous input on interactive clothing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, ACM (New York, NY, USA, 2011), 1313–1322.
15. Kim, D., Hilliges, O., Izadi, S., Butler, A. D., Chen, J., Oikonomidis, I., and Olivier, P. Digits: Freehand 3d interactions anywhere using a wrist-worn gloveless sensor. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, UIST '12, ACM (New York, NY, USA, 2012), 167–176.
16. Lumsden, J., and Brewster, S. A paradigm shift: Alternative interaction techniques for use with mobile & wearable devices. In *Proceedings of the 2003 Conference of the Centre for Advanced Studies on Collaborative Research*, CASCON '03, IBM Press (2003), 197–210.
17. Ng, A., Brewster, S. A., and Williamson, J. H. Investigating the effects of encumbrance on one- and two- handed interactions with mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, ACM (New York, NY, USA, 2014), 1981–1990.
18. Oulasvirta, A., Tamminen, S., Roto, V., and Kuorelahti, J. Interaction in 4-second bursts: The fragmented nature of attentional resources in mobile hci. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, ACM (New York, NY, USA, 2005), 919–928.
19. Schildbach, B., and Rukzio, E. Investigating selection and reading performance on a mobile phone while walking. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '10, ACM (New York, NY, USA, 2010), 93–102.
20. Starner, T. Attention, memory, and wearable interfaces. *Pervasive Computing, IEEE 1*, 4 (Oct 2002), 88–91.
21. Starner, T. Project glass: An extension of the self. *Pervasive Computing, IEEE 12*, 2 (April 2013), 14–16.
22. Weigel, M., Mehta, V., and Steimle, J. More than touch: Understanding how people use skin as an input surface for mobile computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, ACM (New York, NY, USA, 2014), 179–188.
23. Xiao, R., Laput, G., and Harrison, C. Expanding the input expressivity of smartwatches with mechanical pan, twist, tilt and click. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, ACM (New York, NY, USA, 2014), 193–196.

Near-Eye Displays: A Survey on Current Interaction and Application Trends

Dennis Wolf
Ulm University
dennis.wolf@uni-ulm.de

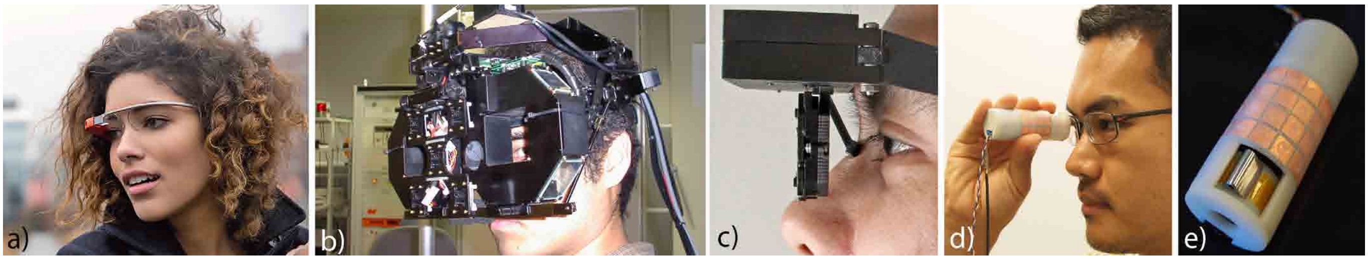


Figure 1. (a) *Google Glass* [40] and (b) *ELMO* (Enhanced see-through display using an LCD panel for Mutual Occlusion) [24] use optical see-through technology, while (c) represents a near-eye light field display [25]. (d) Loupe is used like a spyglass to view a virtual image. (e) It uses touch electrodes for interaction [29].

ABSTRACT
In this paper we introduce current near-eye display technology and give an overview of interaction techniques used in conjunction with these displays. Finally, we summarize current trends in the near-eye research, the technical and social issues that can arise, and discuss the possible development of future near-eye technology.

INTRODUCTION
A Near-eye display or head-mounted display (HMD) is, as its name already implies, a device that is worn on the head and provides its wearer with a display near the eye. An HMD can either be monocular, i.e. provide only one eye with a display like the *Google Glass* [40], or binocular, i.e. offer one display for each eye like the *Oculus Rift* [1]. In addition, these optical displays can be opaque or see-through. On an opaque display only the image on the display itself is visible to the wearer, whilst a see-through display can show an overlay of its own image on top of the real world. This can be achieved by using either optical see-through, i.e. semi-transparent mirrors [24], or video see-through, i.e. combining a video image of the real-world with a computer generated image [44].
Another distinctive property is the positioning of computer generated content. The user interface and information content can either have a fixed position within the wearer's view (head-fixed), e.g. when no head-tracking is available, or have a fixed position within the 3D coordinates of the virtual world so that head movement can change the content's relative position and rotation within the wearer's view (world-fixed) [12].
To present a virtual image to the user, a wide range of display types is available. Apart from the technology to produce pixels, such as liquid crystal displays (LED), cathode ray tubes (CRT), and organic light emitting diodes (OLED), the configuration of the displays themselves can differ between prototypes depending on the requirements. While most HMDs employ one large display for both eyes or a small display for each eye, advanced HMDs combine multiple small displays to increase the field-of-view. In addition, an array of micro-displays and micro-lenses can be applied to create a light field display that can present 3-dimensional images to the user [25].

Due to their immersive nature, HMDs are well-suited for application fields like gaming and simulations. But also other research fields have picked up HMD technology, as we will show in the first part of the related work. One drawback compared to touch-displays and usual monitors is the lack of haptic feedback for direct touch interaction, as the visual content has no physical representation. In addition, classical input devices like a mouse or a keyboard might be not visible to the user, e.g. when using non see-through HMDs, or might be not at hand in a mobile setting. In the second part of the related work, we will discuss possible interaction techniques to counterbalance this limitation and social aspects of these alternative techniques.

RELATED WORK
Near-eye displays have a long history and developed from expensive and heavy artefacts to light-weight and affordable prototypes. This development has lead to a new generation of near-eye research. In the following chapter we give an

7th Seminar on Research Trends in Media Informatics (RTMI '15). February 2015, Ulm University, Ulm, Germany.

overview of the different fields and applications that involve near-eye displays and present interaction techniques that can be applied to control these displays.

Application of Near-Eye Displays

In his visionary article *"The computer of the 21 st century,"* Mark Weiser imagined a computer that is almost invisible to the user and virtual reality that is actually feasible [51]. Now, about 15 years later, we have light-weight HMDs and powerful yet small computers that make his vision seem less like fiction and increasingly like a reachable goal. In the following, we present important research fields that apply near-eye technology and discuss their progress and trend.

Augmented Human

When using a tool like a hammer, we use it like an extension of our own arm. The same way we can extend the capabilities of our own mind, when wearing a head-mounted display. On its own, an HMD is just a means to display visual content, but coupled with the right software, it can augment a human's abilities like the capacity of our memory. Thad Starner, a pioneer of wearable computing, designed a so called remembrance agent that served as an extension of his memory [35]. It can be used to permanently record important events of the user's life and uses user-written notes and current context to present relevant information on a near-eye display at any time. This information can include the biography of the person the user is talking to or a history of the last dialogue the user had with this person. The combination of a private near-eye display and a mobile computing device lets this system feel almost like a natural and always available part of the body.

Of course such a system and every other heads-up display (HUD) needs a proper interface design to increase performance and usability as described by Starner [43]. For example a wearable device should switch from visual feedback to audio interaction when the user is driving a car to complement rather than interfere with the visual driving task. It should also resort to keyboard layouts that are easy to learn for novices but efficient for experts.
The second concern is the privacy issue that arises as soon as a system is capable of collecting and recording data about everything and everyone in its surroundings. Therefore the remembrance agent faces similar issues as private cameras [45]. A misuse of this recorded data could allow hackers to reconstruct a person's day and reveal when a person is leaving the house. This would be a perfect opportunity for robbery and other crimes. Surveillance is also a possible misuse of this collected data and could lead to a digital dictatorship.

Taking his vision even further, Starner presented the *Google Glass Project* in 2013 [40]. This optical see-through HMD is kept very close to the design of ordinary glasses and uses a semi-transparent prism to display information in front of the wearer's right eye as shown in figure 1a. It appears to be floating in the air at a fixed distance of about 3 meters. Via Bluetooth connection to a smartphone, the *Google Glass* has access to the knowledge that can be found in the Internet and can search for relevant information for its user. This allows the users to out-source some of their tasks to their artificial "extensions" of themselves. This could be used to fade in notifications similar to a smartphone or give step-by-step instructions for a task the user is performing right now. Unfortunately, many useful applications are still not available due to the technical limitations of current complex context-recognition.

While the *Google Glass* has made a big step towards integration into everyday use, there are still alternative research trends to improve this integration and increase usability. Reducing the overall weight of an HMD could make the users forget that they are actually wearing them [20]. This would lead to less distraction from the virtual content and provide more comfort when worn continuously.
Additionally, other form factors like a spyglass in figure 1d could be worn around the neck [29]. This way the near-eye display is always available when a quick access to information is needed, but less obtrusive than a permanently worn head mounted display. Its other advantage is the array of touch electrodes that is embedded around the device (see Fig. 1e) and provides interaction while holding the device in front of the eye. A switch on its top can toggle between two focal depths which lets the user place different content on each focal depth. This way a 3-dimensional metaphor allows to view more content despite of a rather small screen size.

Augmented Reality

Analysing a captured video stream, and other means can be used to create an augmented reality on near-eye displays. The resulting overlaid images can be used to present additional information on real-world entities. For example, this can be a triggered reaction to an object in the environment [42]. This way a customer could view prices of other vendors while viewing items in a shop. Road information can be blended onto real roads to reflect road conditions and traffic jams ahead.

A permanent alteration of the view could even simulate visual impairment for designers [5] (see Fig. 2a). This can help to create user-friendly interfaces and experience unfamiliar situations.
As in the previous chapter, alternative form factors have been researched to display augmented reality. Instead of a permanently worn HMD, Rekimoto used a magnifying glass metaphor that allowed users to perceive augmented reality by looking through it [33] (see Fig. 2b).

Another approach is to project computer generated images directly on the retina where it overlays with light coming from the real world. This technology is called a virtual retinal display and is mainly used in military applications [9]. It can adapt to eye-damage and -focus and can provide overlay information like X-Ray images of objects or 3D-models. Thereby it overcomes depth perception issues that usually arise with HMDs [48].

Virtual Reality

Although virtual reality (VR) has a very long tradition, a survey in 1994 showed that it still was not "there yet" [6]. One of the reasons was a lack of immersion that highly depends on which kind of HMD is used. According to Sheedy and Bergstrom, a monocular display allows only partial immersion when compared to a binocular HMD [39]. Despite this knowledge, VR research was restricted to a small number of

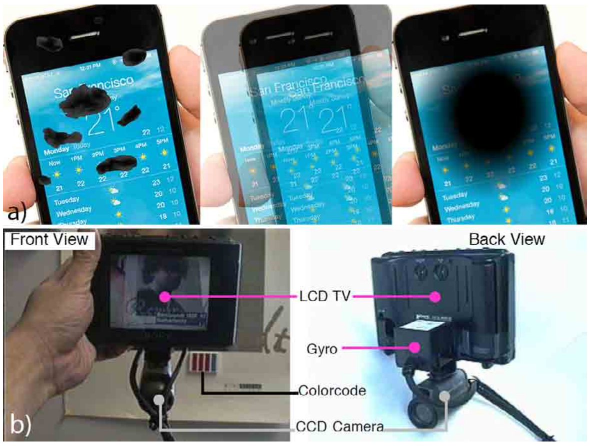


Figure 2. (a) By overlaying and distorting the user's view an experience of visual impairment like Glaucoma can be generated [5]. (b) Real-world images are captured by a camera on the back and presented to user on the front-screen to create a magnifying-glass metaphor [33].

laboratories due to the fact that HMDs were very expensive and uncomfortable. With the appearance of the *Oculus Rift*, the research community now gained a flexible and affordable technology [1]. The main features of the development kit 2 are a large field of view (FOV) of over 90 degrees and fast head-tracking with 6 degrees of freedom (DOF) using a camera positioned in front of the user (see Fig. 3 a). With this new era of HMDs, new application fields have emerged. They include games and simulations to explore dangerous places [11] (see Fig. 3 b), unique experiences of new abilities like flying [34][34] (see Fig. 3 c and d) and teaching scenarios where students can visit 3D scenes of the current topic to increase their motivation [49]. A combination of VR and rehabilitation has also shown promising results in creating a controlled environment for stroke-survivors [18].

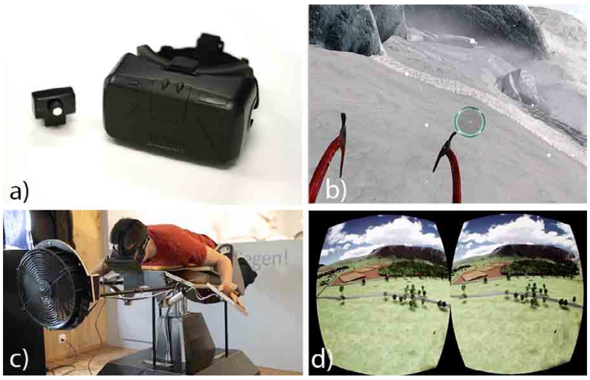


Figure 3. (a) *Oculus Rift* HMD with a camera for head-tracking [1]. (b) *Ascend*, allows the climbing of mountains [11], while *Birdly* simulates a flying experience [34] (c,d).

Interaction with Near-Eye Displays

Due to the proximity of near-eye displays to the human eye, a direct physical interaction with these displays might be uncomfortable. This calls for alternative approaches to interact with the visual content and the hardware producing it. The following chapter introduces interaction techniques that have been either designed for near-eye displays specifically or stem from other application scenarios and can be adjusted for near-eye displays as well. In the following, these interactions will be divided into touch interactions and touch-less interactions.

Touch Interaction

Studies prove how important the right choice of touch-based interactions is for mobile computers [53]. When using touch interaction on the body, the usability and interaction time needed depends strongly on what body parts are touched [50] and what devices are used [4]. These interactions can be used for different purposes.

Text input:
As many applications require text input for queries, log-ins, and other actions, a mobile setup with an HMD is in need of an alternative text input method. This is especially the case when a common keyboard is not available or not visible (e.g. when using an opaque binocular HMD).

In 2004, Lyons and Starner presented the *Twiddler*, a text input device that allowed one-handed and blind text entry [30] (see Fig. 4 a). It uses a chording technique that requires only a small amount of buttons to type letters and special characters. Its small size makes typing inside the pocket possible and is thereby a good solution for setups where the user can not see the actual buttons of the device.

While the *Twiddler* is a device that needs to be hold in the hand for typing, Peshock et al. proposed a keyboard glove that had buttons implanted into a glove for a more comfortable typing experience [31] (see Fig. 4 b). The advantage is that humans can use their kinesthetic memory to hit certain parts of their body, which allows eyes-free interaction with the glove.

Although originally designed for interactions with a smart watch, Funk et al. implemented a text input technique for a wristband that could also be used in combination with HMDs [13] (see Fig. 4 c). Its focus was to prevent occlusion of the actual watch. This is why the touch area takes up only on a small part of the wristband. For the usage with an HMD the touch area can be theoretically expanded to cover the whole wristband. As this technique requires the user to see the symbols on the touch area, it would be recommended to use see-though or monocular HMDs. An alternative would be a virtual augmentation of the wristband to make the symbols visible in video see-through HMDs or VR scenarios with opaque HMDs.

Alternatively, a small wearable QWERTY keyboard could be placed anywhere on the body, where it can be reached by the user [22], e.g. on the forearm like in the study by Thomas et al. [46] (see Fig. 4 d). This way users would be able to use a known keyboard layout in combination with near-eye displays.

Gestures:
Gesture input is another domain of touch interaction. It is often required for navigation purposes and selection of user interface (UI) elements.

Missing any type of interaction device, the *Google Glass* resorted to a touch-field on its frame [40](see Fig. 5a). It

Figure 4. (a) The *Twiddler* uses chording for text input [30]. (b) *Argot*, a wearable one-handed keyboard glove [31]. (c) A touch-sensitive wristband for text entry [13]. (d) A forearm keyboard used in a study by Thomas et al. [22].

supports tapping and stroking gestures as a replacement for missing buttons. The interaction possibilities are very limited in comparison to a full touch-screen but are fully sufficient for the given applications and use cases.

In contrast to worn input devices, Leiber et al. propose the use of tangibles, i.e. real-world objects, to interact with the virtual world [28]. These tangibles can be visualized within the virtual reality and be thereby used by the users to manipulate their perceived content. These objects can also be augmented within the virtual reality to change their form or display content on their surfaces. This way a simple cube can be perceived as a complex multi-surface display.

Focusing on natural and intuitive interaction, Harrison et al. implemented *Skinput*, an input technique that utilizes the user's skin as an input medium. By stretching, tapping, stroking, pinching, and twirling their own skin, users can perform different input gestures as shown in figure 5b. These gestures produce sound waves within the body that can be measured and classified to distinguish gesture types and positions on the skin surface [14].

Similar to *Skinput*, *Nenya* focuses on natural and unobtrusive interaction [3] (see Fig. 5c and 5d). A ring is rigged with a small magnet and placed on a finger, while a sensor is worn on the wrist to measure changes in the magnetic field. By rotating the ring and sliding it along their finger, users can perform different gestures that can be distinguished by their characteristic changes in the magnetic field. This allows unobtrusive gestures like canceling a call during a meeting.

As an alternative to *Skinput*, Serrano et al. propose hand-to-face gestures to interact with devices [38]. While hitting certain face partitions might be a natural task, the social acceptance of these gestures is rather questionable as found by the authors. An additional issue is the sensitivity of facial skin to germs and bacteria which could lead to redness and even rash.

Similar to the use of skin for touch interaction is the use of interactive clothing. *Pinstripe* includes interactive elements into the garment of cloths that recognize gestures [19]. By changing the amount of cloth between the fingers, users can perform distinct pinching and rolling gestures to control devices like an HMD.

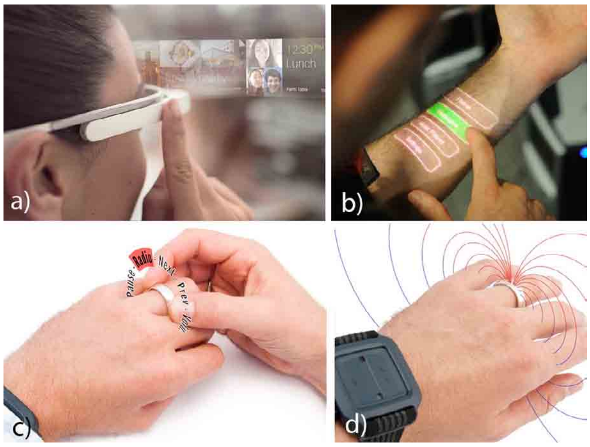


Figure 5. (a) The touch field of *Google Glass* recognizes touch gestures [40]. (b) *Skinput*, using the skin as an input device [14]. (c) *Nenya* allows interaction through rotating and sliding a ring on the finger. (d) The gestures are magnetically tracked [3].

Touch-less Interaction

When no touch is involved in HMD interaction, new means of tracking become necessary.

Hand Gestures:
In 2006, Cheng and Takatsuka proposed a technology for natural interaction with a virtual screen. While using the finger behind the physical near-eye display to select UI elements, users get the impression to interact with the actual virtual content (see Fig. 6a). Therefore the authors refer to this technology as the "virtual touch-screen" [8].

An indirect manipulation technique was proposed by Starner et al. in 2000 [41]. A wearable pendant with infrared-technology was used to track hand gestures and control devices like home-automation (see Fig. 6b). Its advantage is its mobility due to the lack of external tracking.

This tracking approach was also implemented for a watch in 2007 [23]. With this device, a user could perform gestures above the watch and control other devices. This allows eyes-free interaction and can be used in conjunction with both, see-through and opaque HMDs.

As an improvement of this concept, Lee and Starner propose gesture interaction on a wrist-watch that produces vibrations upon a recognized gesture [27] and its successor model that uses a button to confirm a gesture [26]. The advantage of these prototypes is the vibration feedback that allows its use in combination with opaque near-eye displays and even virtual reality. As gestures need to be confirmed by a button press, each action can be undone, i.e. a gesture can be ignored by the user.

A similar approach was proposed by Schindler et al. where an ear-mounted device tracked hand-gestures next to the wearer's head. Its main purpose was topological mapping and indoor localization [37]. Each step taken by a user was recognized through accelerometers and recorded as a path to create a map of an unknown location. Proximity sensors on the ear-worn device detected doorways and labelled these new doorways depending on the finger gestures performed by the user. This way a detailed map could be generated using all collected information.

Eye Gestures:
An unobtrusive and natural way of controlling an HMD is the use of an integrated gaze-tracker [15]. This technology allows hands-free interaction with the HMD and can even be used for text-input. Unfortunately, the light guides used by the gaze tracker to determine the user's focus point in the image lead to a deterioration of the display resolution.

Finger Gestures:
Instead of external movement-tracking, Saponas et al. proposed the use of electromyography (EMG), a technology to measure and classify electrical activity produced by muscles [36] (see Fig. 6c). This approach allows always available gesture-input and the use of novel and natural gestures. EMG technology allows to interact with devices, such as HMDs, even when the hands are occupied by other objects like bags. Providing the user with visual feedback of recognized gestures lead to a higher accuracy but to lower input speed during a study. This trade-off was caused by the users trying to correct their movements until the system correctly recognized all their gestures. Applications with fine-grained controls like the interaction with a machine might profit from this accuracy while non-critical applications like music players might implement the high-speed and error-prone version of this recognition technique.

Similar gestures can also be tracked with *Digits*, a gloveless sensor for 3D-gestures and hand postures [21]. In contrast to the previously mentioned EMG approach, this system uses a vision-based tracking mechanism. Infrared (IR) light is projected from a wrist-worn device onto the user's fingers and captured by an IR camera (see Fig. 6d). The image is then processed to classify the finger posture.

DISCUSSION

Although we have seen many advances in the near-eye technology, there are still technical and social issues to be overcome. A study published in 2002 showed that monocular and see-through HMDs suffered from the rivalry of eyes and inconsistent depth perception [2]. Binocular rivalry of eyes is a phenomenon where different images presented to each eye are not perceived as one overlaid image but alternate between one image in the other instead. A similar phenomenon has also been observed on monocular HMDs where two superimposed images were presented to one eye. Instead of seeing one image, sometimes the clarity of the two images alternated and occasionally one image even disappeared like in binocular rivalry. Additionally, see-through HMDs usually present their virtual images on a fixed focal depth of about 1-2 meters while real-world objects can be present at any distance and therefore have a different focal depth when being focused by the eye. This means that the eyes have to have to refocus between the different focal depth to see a clear image while the other depths are blurry. Usually, our eyes focus on images that crave our attention through changes, so called visual transients. While this has little impact on tasks with a static background, the use of see-through HMDs in combination with a dynamic background or a moving scenario (e.g. driving a vehicle) has been found unsuitable.

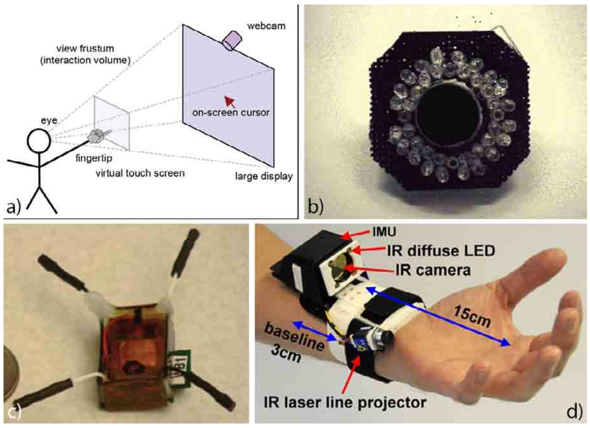


Figure 6. (a) The virtual touch screen allows to interact with a near-eye display in the same way as one would point to a large display [8]. (b) A wearable pendant with IR lights and sensors to recognize gestures [41]. (c) Electromyography requires only a small sensor that needs to be fixed on the arm [36]. (d) *Digits* uses infra-red light to track user hand postures and allow always available interaction [21].

Another difficulty regarding the right choice of near-eye displays is the right field-of-view (FOV). While first HMD prototypes had only a narrow FOV of about 60 degrees, recent HMDs are approaching the natural FOV of a human which is a 180-degree forward-facing horizontal field of view (270 degrees with eye-ball rotation). The benefits of a wider FOV are described in a study of Jones et al., where participants used HMDs with different FOVs to estimate distances in virtual reality [17]. The results show that there is a significant improvement in the estimation of distances with a growing FOV. This might be an indicator that other studies performed with a narrow FOV should be re-evaluated as well.

Concerning interaction with near-eye displays, usability of gestures is often the most important aspect. When these interactions and devices leave the boundaries of a laboratory and become part of our daily life, a new aspect arises. It is the social acceptance. During the evaluation of their hand-to-face gestures, Serrano et al. experienced some "[...]unforeseen results [as] some gestures are considered culturally inappropriate and gender plays a role in selection of specific Hand-to-Face interaction"[38]. This in-acceptance of certain gestures can be culture specific so that a study can produce different results depending on the region it is performed at. To design globally accepted interactions, a lot of effort must be put into the social aspect of this research.

The next social issue is the acceptance of the device itself. HMDs like the *Google Glass* can have embedded video cameras that allow a permanent and secret recording of one's surroundings. While this might be a useful feature for private tasks like recording a car accident or a concert, the recording of other people can lead to privacy violations. As a study of

Denning et al. shows, people feel a discomfort when seeing someone who is wearing a *Google Glass* [10]. In this conflict of interests wearers of HMDs with recording capabilities are sometimes avoided. The fear of being recorded can silence a group of people when such a device enters the room. Although the *Google Glass* has a build-in LED lamp that indicates a recording process so that anyone can see when a recording takes place, this measure is still not sufficient enough to comfort people. This is why special *Google Glass* camera covers are being sold to prominently cover the camera lens. Still, some restaurants have *Glass* ban stickers on their doors to protect the privacy of their customers.

Aside from the acceptance by others, an HMD can have an impact on the wearer as well. Having the possibility to access almost any information at any time, we are less and less depending on our own memory and knowledge. In 2008 Nicholas Carr published an article with the title *"Is Google making us stupid?"* [7]. There he describes how the access to a vast amount of knowledge has lowered his own creativity as an author. One year before that Clive Thompson stated that "[a]llmost without noticing it, we've outsourced important peripheral brain functions to the silicon around us" [47]. The effect of neuronal degeneration that he describes is similar to the out-sourcing of time information. People nowadays do not have the need to guess the current time as they have portable devices like watches and mobile phones to get this information within a few seconds. This accessibility makes it unnecessary to memorize this information, so that usually a second glance is needed if we are asked what time it is. The question here is whether we should out-source these functions to silicon-based intelligence which might be more efficient for this task or train our own brains to stay mentally fit.

CONCLUSION

In this paper, we gave an overview of applications and interactions with near-eye displays. We have shown that head mounted displays have a wide field of application and a large community to push this technology further in its development. These applications are augmented-human abilities, augmented reality and virtual reality. We have also shown different interaction techniques that are either implemented for the use with near-eye displays or can be applied to it. These can be divided into touch interactions like text-input or gestures and touch-less interactions. Taking technical and social aspects into account, HMDs have a bright future as argued by Feng Zhou et al. in 2008 [52].

FUTURE WORK

The limitations of current HMD technologies give us hints on what future trends of research might be. We already mentioned in the discussion that a wider field of view draws closer to the natural view of a human. Apart from higher depth perception there is also an increased immersion which is an important aspect for virtual reality applications. What else can be done to increase immersion? One answer might be wearing comfort. Making the users forget that they actually have a display mounted in front of their faces would reduce the distractions that pull users out of the simulation and back into reality. Except for light-weight glasses and helmets there is also the possibility to mount displays directly onto the user's eye, i.e. on the cornea. A first promising prototype is a contact lens that produces an image with light diodes [16]. It is not ready for application yet as this is just a proof-of-concept. Obvious issues with such displays are the limited power supply and a need of wire-less connection to a computing unit. Future low-energy contact lenses could use solar energy or body warmth to recharge their batteries and have embedded processing units to generate images by themselves. Where would such technology lead our society to? We could be secretly surfing the internet while talking to others or maybe even working on the go. Maybe the human brain would adapt to such a multi-tasked life and learn to process multiple tasks at the same time. Would such a behaviour be socially acceptable or still impolite as today?

What would future interaction techniques with HMDs look like? Maybe gestures will adapt to social conventions, but maybe it will be the other way around as the world evolves around this new technology. Taking it even further, we could measure brainwaves, i.e. electroencephalogram (EEG) activity, to interact with HMDs without an actual movement. Current prototypes are able to measure the strength of brain waves when a person is concentrating on a thought which allows playing simple games like floating a ping pong ball. A more detailed analysis of brain waves would allow complex interactions. The speed of these interactions would be limited by the system's reaction time and not by the user's movement speed. This would allow rapid text-input, fast access of information and would require a completely new generation of user interfaces. Designers would not be limited to the use of certain icon or buttons sizes as they would not be selected by pointers, fingers or cursors. In combination with a natural field-of-view, these would offer a large space for interfaces and information.

Combined with augmented reality (AR), high-level HMDs could allow to place AR content into the real world. A possible application would be a city-, county- or world-wide game that lets players share a virtual game world atop of the real one. A prototype that implemented a campus-wide AR version of the game *Quake* was published in 2002 [32]. It placed foes within the 3D model of the campus that could only be seen through an HMD and be shot with a special toy gun. A more serious application could augment public buildings with reviews of visitors that can be viewed while walking by the building. This way users would have a layer of useful information on top of the real-world and use their HMDs unconsciously.

Augmenting our selves and our environment, near-eye displays will become an essential yet gradually 'invisible' part of our daily life. As we get used to it, interaction with these displays will become as natural and casual as taking a look at one's wrist-watch to check the time.

REFERENCES

1. Oculus VR | oculus rift - virtual reality headset for immersive 3d gaming. `www.oculus.com`, visited 2014-10-31.

2. Rivalry and interference with a head-mounted display. *ACM Trans. Comput.-Hum. Interact. 9*, 3 (Sept. 2002), 238–251.

3. Ashbrook, D., Baudisch, P., and White, S. Nenya: Subtle and eyes-free mobile input with a magnetically-tracked finger ring. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, ACM (New York, NY, USA, 2011), 2043–2046.

4. Ashbrook, D. L., Clawson, J. R., Lyons, K., Starner, T. E., and Patel, N. Quickdraw: the impact of mobility and on-body placement on device access time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2008), 219–222.

5. Ates, H. C., Fiannaca, A., and Folmer, E. Immersive simulation of visual impairments using a wearable see-through display. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility*, ASSETS '14, ACM (New York, NY, USA, 2014), 301–302.

6. Brooks Jr, F. P. What's real about virtual reality? *Computer Graphics and Applications, IEEE 19*, 6 (1999), 16–27.

7. Carr, N. Is google making us stupid? *Yearbook of the National Society for the Study of Education 107*, 2 (2008), 89–94.

8. Cheng, K., and Takatsuka, M. Estimating virtual touchscreen for fingertip interaction with large displays. In *Proceedings of the 18th Australia Conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments*, OZCHI '06, ACM (New York, NY, USA, 2006), 397–400.

9. Chinthammit, W., Seibel, E. J., and Furness, T. A. Unique shared-aperture display with head or target tracking. In *Proceedings of the IEEE Virtual Reality Conference 2002*, VR '02, IEEE Computer Society (Washington, DC, USA, 2002), 235–.

10. Denning, T., Dehlawi, Z., and Kohno, T. In situ with bystanders of augmented reality glasses: Perspectives on recording and privacy-mediating technologies. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, ACM (New York, NY, USA, 2014), 2377–2386.

11. Dufour, T., Pellarrey, V., Chagnon, P., Majdoubi, A., Torregrossa, T., Nachbaur, V., Li, C., Ibarra Cortes, R., Clermont, J., and Dumas, F. Ascent: A first person mountain climbing game on the oculus rift. In *Proceedings of the First ACM SIGCHI Annual Symposium on Computer-human Interaction in Play*, CHI PLAY '14, ACM (New York, NY, USA, 2014), 335–338.

12. Feiner, S., MacIntyre, B., Haupt, M., and Solomon, E. Windows on the world: 2d windows for 3d augmented reality. In *Proceedings of the 6th Annual ACM Symposium on User Interface Software and Technology*, UIST '93, ACM (New York, NY, USA, 1993), 145–155.

13. Funk, M., Sahami, A., Henze, N., and Schmidt, A. Using a touch-sensitive wristband for text entry on smart watches. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '14, ACM (New York, NY, USA, 2014), 2305–2310.

14. Harrison, C., Tan, D., and Morris, D. Skinput: appropriating the body as an input surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2010), 453–462.

15. Järvenpää, T., and Aaltonen, V. Compact near-to-eye display with integrated gaze tracker. In *Photonics Europe*, International Society for Optics and Photonics (2008), 700106–700106.

16. Johnson, R. C. What is electronic contact lens? *SIGDA Newsl. 38*, 2 (Jan. 2008), 1:1–1:1.

17. Jones, J. A., Suma, E. A., Krum, D. M., and Bolas, M. Comparability of narrow and wide field-of-view head-mounted displays for medium-field distance judgments. In *Proceedings of the ACM Symposium on Applied Perception*, SAP '12, ACM (New York, NY, USA, 2012), 119–119.

18. Kaminer, C., LeBras, K., McCall, J., Phan, T., Naud, P., Teodorescu, M., and Kurniawan, S. An immersive physical therapy game for stroke survivors. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility*, ASSETS '14, ACM (New York, NY, USA, 2014), 299–300.

19. Karrer, T., Wittenhagen, M., Lichtschlag, L., Heller, F., and Borchers, J. Pinstripe: Eyes-free continuous input on interactive clothing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, ACM (New York, NY, USA, 2011), 1313–1322.

20. Kasai, I., Tanijiri, Y., Endo, T., and Ueda, H. A forgettable near eye display. In *Wearable Computers, The Fourth International Symposium on*, IEEE (2000), 115–118.

21. Kim, D., Hilliges, O., Izadi, S., Butler, A. D., Chen, J., Oikonomidis, I., and Olivier, P. Digits: Freehand 3d interactions anywhere using a wrist-worn gloveless sensor. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, UIST '12, ACM (New York, NY, USA, 2012), 167–176.

22. Kim, H., Sohn, M., Kim, S., Pak, J., and Lee, W. Button keyboard: A very small keyboard with universal usability for wearable computing. In *Proceedings of the 11th IFIP TC 13 International Conference on Human-computer Interaction*, INTERACT'07, Springer-Verlag (Berlin, Heidelberg, 2007), 343–346.

23. Kim, J., He, J., Lyons, K., and Starner, T. The gesture watch: A wireless contact-free gesture based wrist interface. In *Wearable Computers, 2007 11th IEEE International Symposium on*, IEEE (2007), 15–22.

24. Kiyokawa, K., Billinghurst, M., Campbell, B., and Woods, E. An occlusion-capable optical see-through head mount display for supporting co-located collaboration. In *Proceedings of the 2Nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, ISMAR '03, IEEE Computer Society (Washington, DC, USA, 2003), 133–.

25. Lanman, D., and Luebke, D. Near-eye light field displays. *ACM Trans. Graph. 32*, 6 (Nov. 2013), 220:1–220:10.

26. Lee, S. C., Li, B., and Starner, T. Airtouch: Synchronizing in-air hand gesture and on-body tactile feedback to augment mobile gesture interaction. In *Wearable Computers (ISWC), 2011 15th Annual International Symposium on*, IEEE (2011), 3–10.
27. Lee, S. C., and Starner, T. Mobile gesture interaction using wearable tactile displays. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*, ACM (2009), 3437–3442.
28. Leibe, B., Starner, T., Ribarsky, W., Wartell, Z., Krum, D., Singletary, B., and Hodges, L. The perceptive workbench: Toward spontaneous and natural interaction in semi-immersive virtual environments. In *Virtual Reality, 2000. Proceedings. IEEE*, IEEE (2000), 13–20.
29. Lyons, K., Kim, S. W., Seko, S., Nguyen, D., Desjardins, A., Vidal, M., Dobbelstein, D., and Rubin, J. Loupe: A handheld near-eye display. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, ACM (New York, NY, USA, 2014), 351–354.
30. Lyons, K., Starner, T., Plaisted, D., Fusia, J., Lyons, A., Drew, A., and Looney, E. Twiddler typing: One-handed chording text entry for mobile phones. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (2004), 671–678.
31. Peshock, A., Duvall, J., and Dunne, L. E. Argot: A wearable one-handed keyboard glove. In *Proceedings of the 2014 ACM International Symposium on Wearable Computers: Adjunct Program*, ISWC '14 Adjunct, ACM (New York, NY, USA, 2014), 87–92.
32. Piekarski, W., and Thomas, B. Arquake: the outdoor augmented reality gaming system. *Communications of the ACM 45*, 1 (2002), 36–38.
33. Rekimoto, J. A magnifying glass approach to augmented reality systems. *Presence 6*, 4 (1997), 399–412.
34. Rheiner, M. Birdly an attempt to fly. In *ACM SIGGRAPH 2014 Emerging Technologies*, SIGGRAPH '14, ACM (New York, NY, USA, 2014), 3:1–3:1.
35. Rhodes, B., and Starner, T. Remembrance agent: A continuously running automated information retrieval system. In *The Proceedings of The First International Conference on The Practical Application Of Intelligent Agents and Multi Agent Technology* (1996), 487–495.
36. Saponas, T. S., Tan, D. S., Morris, D., Balakrishnan, R., Turner, J., and Landay, J. A. Enabling always-available input with muscle-computer interfaces. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, ACM (2009), 167–176.
37. Schindler, G., Metzger, C., and Starner, T. A wearable interface for topological mapping and localization in indoor environments. In *Location-and Context-Awareness*. Springer, 2006, 64–73.
38. Serrano, M., Ens, B. M., and Irani, P. P. Exploring the use of hand-to-face input for interacting with head-worn displays. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, ACM (New York, NY, USA, 2014), 3181–3190.
39. Sheedy, J., and Bergstrom, N. Performance and comfort on near-eye computer displays. *Optometry & Vision Science 79*, 5 (2002), 306–312.
40. Starner, T. Project glass: An extension of the self. *Pervasive Computing, IEEE 12*, 2 (2013), 14–16.
41. Starner, T., Auxier, J., Ashbrook, D., and Gandy, M. The gesture pendant: A self-illuminating, wearable, infrared computer vision system for home automation control and medical monitoring. In *Wearable computers, the fourth international symposium on*, IEEE (2000), 87–94.
42. Starner, T., Mann, S., Rhodes, B., Levine, J., Healey, J., Kirsch, D., Picard, R. W., and Pentland, A. Augmented reality through wearable computing. *Presence: Teleoperators and Virtual Environments 6*, 4 (1997), 386–398.
43. Starner, T. E. Attention, memory, and wearable interfaces. *IEEE Pervasive Computing 1*, 4 (Oct. 2002), 88–91.
44. State, A., Keller, K. P., and Fuchs, H. Simulation-based design and rapid prototyping of a parallax-free, orthoscopic video see-through head-mounted display. In *Proceedings of the 4th IEEE/ACM International Symposium on Mixed and Augmented Reality*, ISMAR '05, IEEE Computer Society (Washington, DC, USA, 2005), 28–31.
45. Strub, H., Johnson, K., Allen, A., Bellotti, V., and Starner, T. Privacy, wearable computers, and recording technology. *2012 16th International Symposium on Wearable Computers 0* (1998), 150.
46. Thomas, B., Tyerman, S., and Grimmer, K. Evaluation of text input mechanisms for wearable computers. *Virtual Reality 3*, 3 (1998), 187–199.
47. Thompson, C. Your outboard brain knows all. *Wired Magazine*, 15.10 (2007), 15–10.
48. Van Krevelen, D., and Poelman, R. A survey of augmented reality technologies, applications and limitations. *International Journal of Virtual Reality 9*, 2 (2010), 1.
49. Villagrasa, S., Fonseca, D., and Durán, J. Teaching case: Applying gamification techniques and virtual reality for learning building engineering 3d arts. In *Proceedings of the Second International Conference on Technological Ecosystems for Enhancing Multiculturality*, TEEM '14, ACM (New York, NY, USA, 2014), 171–177.
50. Wagner, J., Nancel, M., Gustafson, S. G., Huot, S., and Mackay, W. E. Body-centric design space for multi-surface interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, ACM (New York, NY, USA, 2013), 1299–1308.
51. Weiser, M. The computer for the 21 st century. *ACM SIGMOBILE mobile computing and communications review 3*, 3 (1999), 3–11.
52. Zhou, F., Duh, H. B.-L., and Billinghurst, M. Trends in augmented reality tracking, interaction and display: A review of ten years of ismar. In *Proceedings of the 7th*

IEEE/ACM International Symposium on Mixed and Augmented Reality, ISMAR '08, IEEE Computer Society (Washington, DC, USA, 2008), 193–202.

53. Zucco, J. E., Thomas, B. H., Grimmer-Somers, K., and Cockburn, A. A comparison of menu configurations and pointing devices for use with wearable computers while mobile and stationary. In *Proceedings of the 2009 International Symposium on Wearable Computers*, ISWC '09, IEEE Computer Society (Washington, DC, USA, 2009), 63–70.

Non-Contact Actuation of Matter in HCI

Evgeny Stemasov
University of Ulm
evgeny.stemasov@uni-ulm.de

ABSTRACT
In the following paper, three concepts for contactless actuation are going to be presented and compared: ultrasound, including standing and travelling waves, magnetic fields and actuation by air vortices or jets. For every method, two general use-cases are possible: the actuation of tangible objects, to allow a computer to communicate with a user on a physical layer and the actuation of human users themselves, which allows the enhancement of gestures with tactile feedback. Firstly, physical basics, variations and implementations of each concept will be presented. Afterwards, problems and challenges in an HCI-context will be shown. Finally, the general methods are going to be compared, as they differ in aspects of mobility, precision and bring different limitations with them. Based on this, an outlook on conceivable research directions will be given.

Author Keywords
Non-Contact Feedback, Remote Feedback, Tangible Interfaces, Haptic Feedback, Human-Computer-Interaction

ACM Classification Keywords
H.5 Information Interfaces and Presentation

INTRODUCTION
Most modern interfaces tend to communicate with the user by only using visual and audible channels - the sense of touch, a humans most intricate sense and mean of perception is usually ignored or barely demanded. This is lost potential, as studies have shown that proper haptic feedback can reduce error rates and improve the user experience [36, 5, 22]. Another problem that arose from the introduction of tangible interaction objects, is that while it is possible for the user to manipulate these objects, the system he is communicating with can only perceive these objects. On this haptic layer, interaction seems to be a one-way-street. In the following paper, multiple concepts for making this communication bilateral will be presented - meaning how to remotely actuate matter, may it be the user for haptic feedback or objects in the user's vicinity.

There are three general concepts: actuation by ultrasound, including standing and travelling waves, by electromagnetic fields and air jets or vortex rings, each of them has its own advantages and disadvantages. All these techniques can be used to expand and improve existing interfaces like VR/AR Systems and gesture-based channels. In comparison to interfaces that require some kind of tethering like cables or gloves, interfaces employing remote feedback can provide it without significant alterations of the user and therefore can be used in a "walk-up-and-use" context - a core factor in current research.

METHOD 1: ACTUATION BY ULTRASOUND

Introduction and Physical Aspects
The phased ultrasound array is responsible for the generation of acoustic waves. It consists of generic ultrasound transducers, which are commonly used in parking sensors [4]. Multiple focal points are achieved by either grouping parts of the array or by temporal multiplexing. Grouping leads to a weaker applied force as less transducers are attributed to one focal point, while temporal multiplexing leads to a slightly discontinuous force exertion. A focal line can be generated, when the setup consists of 4 arrays, by generating multiple focal points one after another. The phenomenon used to generate tactile feedback with ultrasound is called "acoustic radiation pressure" - when an object is in the way of the ultrasound-propagation, a pressure field is created on impact by the fact that most of the ultrasound is being reflected [11]. This phenomenon is the key to the systems presented in the first part of this paper. For humans, 99.9% of the acoustic energy are reflected at the skin surface, the remaining 0.01% are considered harmless [14].

Perception of Ultrasound
For the perception of such impulses, in the human hand two receptors are responsible: the Meissner corpuscle which is unevenly distributed around the area of the hand, reacting to low-frequency vibration and the Pacinian corpuscle which is evenly distributed and reacts to high-frequency vibration. The fibre around both of them is rapidly-adapting, meaning that it detects the transients of skin deformation and discharges at the beginning and the end of a mechanical stimulus [37]. Pacinian corpuscles usually cover the whole fingertip and are also evenly distributed across the whole hand, while the Meissner corpuscles have a lower density at the fingertips [37, 20]. Their function covers feedback signals required for grip control of an object - e.g. the recognition of a slipping item held in a hand, which is generating "low-frequency skin-motions" [37].

7th Seminar on Research Trends in Media Informatics (RTMI '15).
February 2015, Ulm University, Ulm, Germany.

Earlier studies have shown that the accuracy for localizing a single point of stimulation by ultrasound on a finger is rather low, at 50-60% [37, 4] - this value does not change significantly, even after days of training [37]. Over 60% of stimulation points are even recognized on the wrong finger [37]. When receiving ultrasound stimulation on the palm of the hand, drifting set in and many subjects started to localize the impacts closer to the thumb or the wrist than they actually were [37]. The conclusions of most studies have been that the shorter a stimulus, the worse movement can be recognized [37, 9]. A minimum duration of an impulse is about 50 ms to give the user a good feeling of movement and the more points between the beginning of a simulated movement and its endpoint, the better the transit is recognized as such [37].

UltraHaptics

The UltraHaptics system has been developed in Bristol and its goal is to provide haptic feedback above an interactive surface [4]. The phased array of this system employs transducers in a 16 x 20 Grid. Hand tracking is implemented through a LeapMotion controller, visual output is handled by a projector above the active area, while the transducer array is placed behind the projection surface [4]. This requires the material of the screen to be an "acoustically transparent display" [4], which essentially means that the ultrasound can pass through it mostly unhindered and with minuscule refraction. A plane with holes or other regular structures affects the quality of the projection negatively, it distorts the image or starts to show moiré-patterns - effects which both decrease the user experience [4]. The optimal material has to be solid and even for best projection quality and also acoustically transparent to at least one frequency, to allow proper transmission of the ultrasound through it. The researchers tested multiple materials for the attenuation of ultrasound and its diffraction when passing through the screen and concluded a significant tradeoff between permeability to ultrasound of a certain frequency, meaning how much open space the material offers and the quality of the projected image [4]. It has also been concluded, that smaller holes in the material reduce diffraction more than the total open space a plane offers [4].

UltraHaptics can be used for multiple applications in mid-air/above the screen. Carter et al. present a setup of an image viewer with a mid-air pinch-to-zoom functionality. One focal point is generated below the thumb and another at the active finger. When the user moves the fingers apart, the frequency difference of the two points grows and weakens when the user pinches or extends his fingers respectively [4]. A simple media player has also been implemented - a focal point signifies the play/pause button, another the volume slider. The latter can be grabbed and moved by the user. The strength of these focal points differs and therefore, after a training phase, the user is able to discern these controls blindly as he is "being guided" to the interface elements [4].

HaptoMime

HaptoMime has been developed in Tokyo and presented at the UIST in October 2014. It combines the technology of a floating holographic screen with the aforementioned ultrasound feedback concept. The floating-screen-technology, brought to industrial readiness by the Japanese company ASUKANET [24] among others is primarily based on the so-called aerial imaging plate (AIP), which reflects an image from an ordinary display to a symmetric position in front of the AIP, where it appears to float in front of the user, depending on his position [20]. The problem with such floating imagery is that is "inherently lacks tactile feedback" [20]: when interacting with a hologram-like interface, the user tends to insert his finger or stylus way beyond the visual limit of buttons and the like [20] - this can be solved by providing tactile feedback for the user. However, a problem arises: it is not possible to place the phased ultrasound array behind an LCD, as it is not acoustically transparent. It is not possible to put an LCD behind the array either, as it is not visually transparent. This obstacle is bypassed with an elaborate transmission and reflection scheme, which can be seen in Figure 1. Ultrasound can be reflected by the AIP, which is placed between the LCD and the phased array in a 45 degree angle. The image of the screen floats in front of the AIP, mirrored at the plane of it [20]. Sensing is implemented through an infrared frame around the plane of the floating screen. According to the input of this sensor, the ultrasound phased array delivers focal points to the user's fingertip [20].

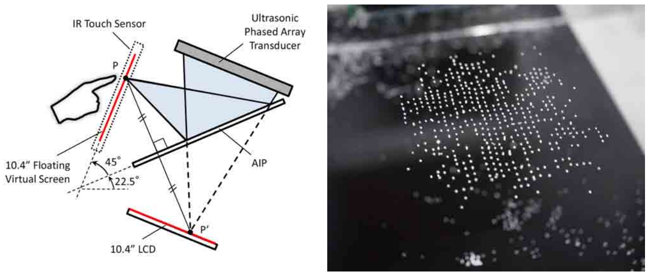


Figure 1. Simplified schematic setup of HaptoMime as presented in [20] (left), levitating particles in an acoustic potential field as implemented in Pixie Dust [23] (right)

A bigger setup, which essentially follows the same scheme has been presented by Inoue et al. [12]: they employ the same kind of aerial imaging plate, but offer a bigger interaction space with 4 ultrasound transducer arrays around a cube of 35 x 30 x 27 cm [12]. In this concept, the LCD is mounted on a dynamically adjustable stage to change the depth of the image to the user's preference [12]. A key difference is that it is not required to redirect the ultrasound, as the transducers are placed at the sides of the interaction space, while the aerial imaging plate is placed at the back side of the cube [12].

There are three technical demonstrations for HaptoMime, evaluating interaction concepts with a tangible floating display: a calculator and a small piano with 12 keys demonstrate the possibilities of floating, button-based interfaces, which do not require any physical touch, yet provide haptic feedback. The third demonstration shows a simple drawing application, which provides a friction or resistance effect when the user is dragging his or her finger on the canvas [20].

Pixie Dust

The two projects just presented utilize ultrasound to generate a pushing force on the user's finger, Pixie Dust by Ochiai et al. chooses another approach with ultrasound: it uses the phenomenon of acoustic levitation discovered and patented in the eighties [29] to keep small and light objects floating in mid-air - this allows remote, contactless actuation of objects. To support objects against the force of gravity, standing waves are generated by four ultrasound phased arrays around the active area. The nodes of these standing waves can then support small objects - by altering the standing waves and nodes, movement and animation becomes possible [23]. The setup consists of 4 transducer arrays, driven by an FPGA board. Sensing, if necessary, is implemented through IR-cameras or alternatively by a Kinect to sense the levitating objects and user input [23]. For most demonstrations, small Styrofoam or polystyrene spheres with a diameter of 1 mm [23] are used, as they are easier to manage, due to their symmetric shape and low weight. The first application is a projection screen - its resolution is 85 x 85 particles, aligned in intervals of 4,25 mm. The particles can be moved parallel to the screen to achieve a seemingly higher resolution, while movement orthogonally to the screen plane achieves a volumetric effect [23]. To use Pixie Dust for levitated raster graphics, all nodes in a plane are filled with particles, afterwards, an air jet or an additional phased array blows away excess ones to create a raster image which can, if needed, be extended and illuminated with a projection [23]. Vector graphics are displayed by calculating paths for single or multiple particles to follow and animating them accordingly. They move with a speed of 72 cm/s which is enough for the human eye to perceive some kind of consistent image [23].

Related Work

UltraTangibles, also developed by Subramanian et al. allows actuation of tangible objects on a small tabletop interface. Key idea is the coupling of digital content with its real-world representation, which usually is unidirectional: a user can move a tangible for the system to sense and react, but the system's reaction can not move the physical representation of its internal model [19]. UltraTangibles employ four ultrasound phased arrays around a 7 inch display to actuate small plastic spheres. In total, 144 transducers are used: two 15 x 3 arrays at the longer sides of the rectangle and two 9 x 3 arrays at the short sides. Sensing is implemented through a Playstation Eye used in the from the PS3, which provides up to 100 fps [19] as input for a PC driving the circuit boards of the transducer array. When multiple objects are involved, temporal multiplexing or splitting of the arrays becomes necessary, as mentined before. For movement, an initial pulse is fired, the resulting displacement is monitored and corrected by smaller pulses [19].

Demonstrations involve for example a control scheme for a small sphere which is moved to a location the user indicates by touching the screen - up to two spheres are supported in this setup [19]. Another demo is a game, where two digital paddles play the well-known pong-game with a tangible ball [19]. Combined with the sensing input, it is also possible to record and replay interactions the user has with the tangibles or save and recall old positions or configurations of the objects [19]. A possible extension is another array below the screen, if it is "acoustically transparent" or a change to projection as visual output: this would allow the alteration of friction forces and the potential use of heavier objects as tangibles [19].

Touchable Holography uses a holographic display [11] - which essentially provides a similar effect to the AIP presented before - and one phased array mounted above the user's hand. It provides one single focal point which is dynamically movable [11]. Tracking is implemented through a WiiMote and a reflective marker attached to the user's middle finger [11]. Demonstrations were a Rain-Demo, where the user can feel drops falling on his palm and a small creature walking across his hand. Both use-cases provide consistency between digital representation and physical/tactile experience of the user. Similar approaches were used in various other papers, like the "Noncontact Display" [9], the "Airborne ultrasound Display" [14], compact versions [8, 10] and a bigger setup in [7]. The rendering and recognition of virtual objects has been dealt with by Long et al. in [18].

METHOD 2: ACTUATION BY AIR

Introduction and Physical Aspects

A medium available nearly everywhere is air, however, it is incredibly hard do control and therefore poses a challenge for researchers attempting to use it in an HCI-context. As it is the case with ultrasound, there are two general modes of operation: actuation of objects, as shown in [1, 13, 30] and feedback for human users as can be seen in [31, 5, 30]. There are also two types of actuation: Vortex rings, generated by speakers pushing air through a circular hole and air jets, consistently pushing air out of a nozzle. Air vortices are more stable and can travel up to 7 meters [5], though they are rather vulnerable to ambient airflow coming from air draught, open windows, air conditioners or even the movement of users in the room [1, 30]. Air jets suffer from strong dispersion - after travelling a certain distance, where the feedback is constant and rather precise (Region I in Figure 2), the flow starts to diverge at a degree of about 14°, depending on the nozzle shape [13] (Region II in Figure 2). Generating an air vortex has already been described in 1867 [5] - a sheet pushing air out of a circular opening. Nowadays, the membrane is an electronically driven speaker in both vortex-based projects. (Sodhi et al.: AIREAL [30] and Gupta et al.: AirWave [5]) Basic research on vortex speed, stability and force has been conducted in [5], with the results being that there are heavy tradeoffs: Stable and large vortices travel big distances, but apply a "barely perceptible" force [5].

Air Vortices

Air vortices have been used and evaluated in two projects: AIREAL [30] and AirWave [5]. AirWave is a research-oriented vortex generator consisting of one speaker in an acrylic glass case on a manual pan-and-tilt platform, with a laser sight for aiming [5]. Different, swappable faceplates to adjust size and structure of the vortices were used to analyze their perception in terms of impact strength, precision, "fluffiness", size and hardness. There are also other modes of operation for this setup, as it is possible to use perfume for "smell-feedback", aiming at the users face with big, stable but also

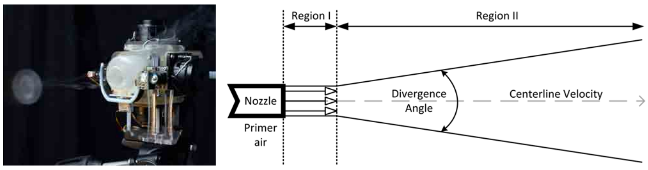


Figure 2. Air vortex as generated by AIREAL [30], filled with smoke (left) and scheme of the flow regions, as described in [32],[33](right)

weak vortices. Smoke may be used not only for visualizing the vortices for research purposes, but for expanding haptic feedback to a visual layer, too. Another variation is the use of hot or cold air in combination with humid or arid air, which has not been thoroughly researched yet, but already used for entertainment purposes together with olfactory in so-called "4D-Cinemas" [21]. For demonstration, only augmented gestures in front of a tablet were presented [5] - however, if an automatic aiming system is added, most uses shown by the AIREAL-System are possible.

AIREAL has been developed by Disney research and essentially provides an aim-able vortex generator as presented before. A 3D-printed enclosure with 5 15W subwoofers, together with a flexible nozzle for aiming, provides a small, mobile vortex generator to be mounted on tripods or interactive surfaces [30]. The AIREAL system can be used in various setups: 1) mounted on one Tripod and aimed at the user or his environment 2) Two generators above a TV for haptic feedback 3) One generator beside a tablet to augment mid-air gestures 4) Haptic projection: meaning that the system combines projections (e.g a butterfly) with the corresponding and expected haptic feedback [30]. Actuation is not limited to human skin, as the butterfly-projection has also been demonstrated on a plant, which shook to the insect's movement [30].

Air Jets

Air jets are a simpler form of air feedback: a compressor pushes air out of a nozzle to generate a constant airflow. Early attempts (e.g [31]) provided an array of nozzles in a tabletop controlled by special valves and required the user to carry a spoon-like tool to "catch" the airflow and feel its force accordingly [31]. The simulation uses the first air flow region (Figure 2) to provide force feedback, as it shows more predictable behaviour, due to the divergence setting in later. In combination with a stereoscopic projection, one can see and feel virtual shapes on a table [31]. Unlike interfaces like this VR-extension presented in 2005 by Suzuki and Kobayashi [31], Tsalamlal et al. [32, 33] used the second air flow region (Figure 2) for user feedback, which demonstrates high diffusivity and bigger area of impact, depending on the distance travelled [32]. The setup consists of a robotic arm which moves an air jet pointing upwards along a flat surface, above this tabletop a Kinect is used to track the user's hand and alter the nozzle's position accordingly [32].

METHOD 3: ACTUATION BY ELECTROMAGNETIC FIELDS

Madgets

Madgets is a system of tangible tabletop widgets developed in Aachen [35, 34]. A 24 inch TFT-display serves as an active surface, an electromagnet array below is responsible for actuating widgets on the surface, by dynamically switching

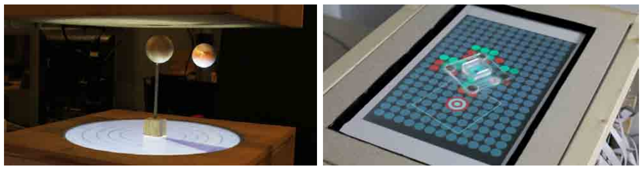


Figure 3. The planetarium demonstration from ZeroN [17], (left) and a button Madget moving towards a target [34, 35] - where the magnet polarization is colour-coded accordingly. (right)

their polarity. Visual tracking is implemented through multiple cameras below the array, while a matrix of polymer fiber-optic cables serves as a low-resolution multitouch sensor [35]. There is a multitude of "passive Madgets" - passive meaning here that the tangible can only be moved, however, there are also active Madgets, like actuated radio-buttons or dials, where the user and the system can alter the state of the control elements [35] - a novel progression. It is also possible to harvest energy for functions in Madgets through induction - an application demonstrated an LED glowing without being directly tethered to a power source [35]. Most movement happens in 2D-space, while there are small extensions to that constraint: by using magnetic repulsion to propel a magnet upwards, a bell can be actuated for system feedback or to alter the state of push-buttons [35].

FingerFlux

FingerFlux provides a system for near-surface haptic feedback on tabletops. On a grid of 19 x 12 electromagnets, each magnet is individually controlled by a software to create an attracting or repulsing force to a magnet attached to the user's finger [36]. Attraction and repulsion can be combined for different applications: FingerFlux can limit user movement to a line above the surface, using a line of magnets set to attraction and a frame of them around it set to repulsion. It can also be used as a pointing augmentation, comparable to 'edge-snap' in image editing tools. As an addition, it is possible to vibrate the user's fingertip above the surface, by quickly switching polarities. The maximum distance to the tabletop is on average 35 mm, while more sensitive users felt vibrations and other forces at a higher distance [36].

ZeroN

ZeroN provides a setup for magnetic levitation of a sphere in 3D-space. A strong electromagnet is placed above the interaction space and can be moved freely. To sense the distance of the sphere to the levitator, a Hall-Effect sensor is included [17]. To track the object, 2 cameras are used, while user actions are sensed with a Kinect. A projector augments the interaction space with shadows and textures [17]. Multiple demonstrations have been implemented: One can use the sphere as sun to simulate different illumination scenarios or use it as a tangible input for the camera position in CAD-Tools; there is also a small planetarium with one static planet and a moving moon - the path of the satellite is automatically calculated by the position and the mass of it [17].

USES IN HCI AND RESULTING PROBLEMS

Actuation Through Ultrasound

As mentioned before, ultrasound actuation offers 2 modes of operation: Moving objects and creating a touchable surface in mid-air - the hardware required, meaning the phased array stays the same, which would potentially allow dual-use. However, the surface being actuated has to reflect ultrasound - therefore, absorbing materials like thicker clothing are less suitable than human skin or polymers, as they either affect the mechanoreceptor's perception or weaken the exerted force [7, 22]. The limitations are even bigger for levitation like in [23], as additionally, shape and weight have to be considered.

When dealing with objects like the spheres in UltraTangibles it is problematic to stop moving objects - they require strong correctional pulses [19] and tend to "overshoot" slightly. Managing multiple objects - meaning more than two tends to be rather complicated: as mentioned before, one either splits the array in multiple parts, or multiplexes the entire array [19].

Not every surface is suitable for letting enough ultrasound pass through - a limitation to projection arises, unless the arrays are placed around the screen - this in turn suffers from occlusion by objects or impractical ways to interact, due to physical constraints. If a virtual surface for the user to touch is generated, it is still not solid - it is a resistance, which can be passed, therefore it may take time for the user to adapt to this entirely new feeling. Studies conducted have shown that object recognition, frequency differentiation and focal point search get better with training [18, 9].

Actuation Through Air

As the stroke model [30] provides usable formulas to use for stable vortices, the target now is to find an optimum between applied force and vortex stability (and therefore feedback precision). Different nozzle shapes alter the results significantly - a move-able tip, for instance, weakens the vortex notably [5] but increases the speed of aiming and feedback, considering that moving the whole enclosure would take more time and affect vortex generation. AIREAL provides 100% accuracy at 0.5m, 84% at 1.25m [30] - this has to be compared to the use-cases intended: Augmentation of gestures above a tablet happens at a distance of approximately 30-50 cm - for this use-case the precision is sufficient, for multi-user setups and bigger environmental configurations, e.g an immersive gaming environment, not every pulse will arrive on target and therefore may affect the user experience negatively.

Related to this problem is the issue of perception: vortices aimed at the user's shoulders often were perceived on the cheeks [5] - with a targeting resolution of approximately 10 cm it is also unreliable to differentiate between certain facial regions (e.g. left vs. right cheek) [5]. A limitation is also the fact that one cannot predict the way a user is clothed and the perception of an air vortex is heavily impacted when it has to pass multiple layers of clothing of varying thickness [5]. In a study for example, a user wearing a fleece pullover did not react to a vortex at all, even if it visibly shook the clothing [5].

Considering the multitude of concepts presented in [30], air vortex feedback appears to be the most dynamic and free concept in terms of ease of setup. Small vortex generators as the ones in AIREAL, can be easily mounted on tripods, around or above screens or even on dynamically moving mechanical arms like in [32]. This also implies that these kinds of systems are quick to set up and reconfigure.

When actuating objects with air jets, multiple problems arise: The actuation can only happen in a direction away from the nozzle, meaning that for now, it is unidirectional - techniques for suction (Like in VacuumTouch [6]) are not strong enough to generate a force sufficient to actuate objects or provide feedback for the user at a distance. Yet, it is still possible to actuate objects in multiple directions, by either using two or more air jets as intended in [13] or by using gravity as the counterforce as shown in AerialTunes [1]. Iwaki et al. proposed a technology for actuating a cylindrical object on a plane surface - which essentially limits movement to two dimensions and a specific object shape, in this case, a cylindrical one [13]. Three-dimensional movement by air jets has not been implemented in a viable way, as it would require a high computation effort and would be very susceptible to the smallest ambient air flows.

Unlike Vortex Rings, simple air jets suffer from a high diffusivity. While some projects utilize it ([32, 33]), others consider it a big problem ([31]). Depending on the nozzle shape, the angle of expansion is about 14° [13], which, at greater distances implies an even bigger actuated area than, for example, AirWave produces.

Probably the biggest problem air-based actuation methods are facing is the delay [30, 5, 32]: A Vortex ring or a sudden change in the airflow take, depending on the pressure they are generated with and the distance they have to travel, several hundreds of milliseconds to reach their target, including the computation time. AIREAL, for instance takes has an average latency of 139 ms [30]. This does not yet consider any computations that have to be done for aiming, which would involve processing visual input and calculation of output parameters, in this case the vortex target. HAIR suffered from the same problem, which they attributed to the robotic-kit they were using [32]: when testing the time a user takes to find multiple objects on the virtual surface, the result was that all contact-based feedback methods yielded better results, due to the fact that their feedback delay was lower [32].

Actuation Trough Electromagnetic Fields

When comparing the magnetism-based projects presented, it is important to differentiate between two-dimensional and three-dimensional movement capabilities. ZeroN freely moves objects along three axes in space, while Madgets can only be moved in two-dimensional space with small extensions into the depth layer, like the bell widget [35]. When it comes to actuation of objects, magnetic fields seem to provide the steadiest and kind of movement, it can not be considered smooth though: The Planet in the ZeroN simulation tends to shiver when moving or when touched by the user [17]. The animation provided by Madgets also suffers from irregularities, due to the sequential approach of the magnet controls [35].

A downside of magnetism is the limitation of the possible materials - actuated objects are made of magnetic substances or at least a core/surface thereof is required. To allow virtual rotation, ZeroN for example used a metal ball with a disconnected and freely tunable plastic shell, which is tracked visually, as rotating a dipole magnet made it fall to the ground [17]. Another downside to magnetic actuation of objects is that it is only possible to hover and move one single object in a magnetic field. Systems like Madgets, consisting of an array of magnets are able to manage multiple objects, as one magnet in the array is attributed to exclusively one on a tangible and is responsible for moving it until it leaves its active region, which essentially allows a big number of tangibles on the surface, limited primarily by the area and the available computation capabilities.

Systems like FingerFlux require a magnetic attachment for the user to wear [36], as the human body is not magnetic - this limits the possibilities in a "walk-up-and-use" context - potential users would have to put on a magnet before interacting with the system. Tough, it synergizes incredibly well with magnetic implants - a trend in the "biohacking"-community. Allegedly dozens of people [25] implanted small neodymium magnets in their ring finger, to be able to feel electromagnetic fields around them, for instance generated by subways or microwaves and to be able to pick up small magnetic objects. While unlikely to make it to the mainstream, this would allow a use of magnetic human-computer-interfaces without attachments and an allegedly "more natural" feeling [2], as the magnetic implant does not appear to the user as an disconnected part in his finger, but his very own body vibrating and being pulled somewhere [2]. These kinds of invasive workarounds most likely will not be picked up by the target audience in the near future, as there are not only risks of the magnet corroding, but also the inability to receive MRT scans [2] that make the implant unattractive.

With FingerFlux, blind use of interfaces becomes easier as drifting is reduced [36] - being similar to edge-snapping in various image-editing software, it would give hints to the user where a clickable area is and where there are none or unavailable interface elements are. It has to be considered that strong magnetic fields, unlike air vortices or ultrasound may pose a risk to people: not only classical hard drives and magnet tapes but also pacemakers can suffer from magnetism. The range of remote actuation ranges from 0-15 cm [36], if precise movement and targeting is required, more or stronger magnets are necessary [15].

Perception and Reactions

Being a rather novel technique, actuated interfaces are not yet well researched in terms of user acceptance and reaction - most research focuses on technical possibilities, instead of conceptual and perceptional problems [1]. Rasmussen et al. [1, 28, 27] conducted extensive user studies with 3 of their projects: TurningTV, Aerial Tunes and the coMotion Bench. Aerial Tunes involves remote actuation by air jets: the setup consists of 5 boxes generating a vertical airflow, in which a sphere is balanced and held at a certain height. Users can walk up and manipulate this sphere to re-position it in the airstream or remove it entirely - the position in turn alters an ambient soundscape. The reception was positive, while the interpretations and approaches people showed, differed a lot [1]. A system seemingly ignoring the laws of physics has been proven to work as an "eye-catcher" [27], inducing curiosity in people passing by [1] and inviting to explore its functionality. It has been concluded, that people tend to attribute animal or human qualities and behaviours to actions of machines. For example, when the TurningTV followed one of the study participants around the room, others considered the TV to "be in love" with its target [27], while it certainly is not possible for an inanimate object to feel love. This is a criterion to be considered when designing interaction concepts in actuated interfaces - especially when the user is unprepared and untrained - and has not been researched too well, yet.

FUTURE WORK

For future work, a multitude of possible concepts and directions appear:

1. Air Jets do not support proper three-dimensional movement yet - a combination between the concepts presented in AerialTunes [1] and by Iwaki et al. [13] seems to be viable at first glance. Tough, problems with computation and speed of sensing may arise, affecting stability. This would be particularly problematic when it comes to applications like the ones presented in ZeroN [17] where slow but steady movement is required (e.g. for applications like the planetarium demo).
2. As proposed in [15] by Karunanayaka et al., the resolution of near-surface feedback systems can be improved by increasing the density of the magnet units. This means that the resolution for the feedback provided can still be increased. However, computation time and efficient sensing methods have to be considered here. There may be also technical constraints concerning the minimum size of an electromagnet with switchable polarity.
3. Miniaturization of phased arrays for mobile use as proposed by Carter et al. in [37] would allow the augmentation of gestures around/near mobile devices, as long as pocket-sized arrays consisting of smaller transducers can generate sufficient force without appearing bulky or impractical.
4. Palpable health risks have not been properly evaluated for magnetism-based interfaces yet - while magnetic implants pose the absolute minority and can be safely ignored in research for now, pacemakers can be negatively affected by such systems - a risk that should be evaluated soon. Another concern are hard drives, as especially for tabletop systems like Madgets and FingerFlux [35, 34, 36], uses in combination with "normal" computers are conceivable.
5. Air vortex-based actuation methods can not yet provide constant force exertion, they cover only pulse-wise actuation - with sufficient generation frequencies by one or multiple vortex generators, systems like AIREAL [30] may surpass the noticeable difference between vortices and create an illusion of continuity.
6. Not only gestures and environments can be augmented with remote actuation, but also holographic screens as analysed by I. Rakkolainen in [26] or other mid-air visualization technologies as used in HaptoMime [20] or HORN [12] - haptic feedback and the ability to touch what is presented, may improve the user experience.

7. Most research focuses on the actuation of solid objects - however, liquids may probably also be actuated with two of the techniques presented in this paper: Long et al. [18] for example used oil to visualize ultrasound focal points. Furthermore, there are magnetic liquids [3] which may also be actuated by systems like ZeroN or Madgets. Their actuation may be much harder to implement, as a liquid attracted magnetically behaves differently than a solid object [16].

CONCLUSION

In general terms, the main advantage of remotely actuated interfaces is their liberty, when it comes to the use context: it is usually not necessary to change anything about the user to interact with the system, in a VR-Context, one would have to consider that it is not a necessity to leave the user completely untethered, participants are wearing glasses or other attachments anyway, so a "walk-up-an-use" scenario becomes less relevant. Though, it is an important research direction to minimize alterations a user has to make before using a certain interface.

When it comes to actuation of objects, a multitude of possible use-cases in tabletop-systems arise, as these technologies enable bilateral communication between human and computer on a physical layer. While there are limitations in terms of shape, materials and composition of such tangibles, usable setups already exist and can be used as a base for future research and consumer electronics already.

Ultrasound-based systems excel in terms of precision and resolution, in comparison to magnetic or air-based setups. However, they lack range and are relatively costly to set up if bigger numbers of transducers are involved as in [7]. When it comes to range, air vortex based systems are preferable, though, they are most vulnerable to external influences - an opening door can already affect the targeting. Magnet-based systems offer stable levitation and a comparably high resolution at a lower range and donate themselves especially for near-surface actuation. Using air vortices, it is not possible to exert constant force, but only to actuate pulse-wise - air jets provide constant force, but suffer from high diffusivity at greater distances and are therefore hard to control and aim. A somewhat natural feeling for the user and a certain degree of immersion is offered by every concept, even if no method can properly create absolutely "believable matter" as such - however, the discrepancy between the visual and the haptic perception of virtual objects is significantly lowered.

REFERENCES

1. Alrøe, T., Grann, J., Grönvall, E., Petersen, M. G., and Rasmussen, J. L. Aerial tunes: Exploring interaction qualities of mid-air displays. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, NordiCHI '12, ACM (New York, NY, USA, 2012), 514–523.
2. Berg, D. Body Hacking: My Magnetic Implant. `http://www.iamdann.com/2012/03/21/my-magnet-implant-body-modification`. [Online; accessed 10.12.2014].
3. Campbell, M. Magnetic Field Sculptures. `http://www.newscientist.com/blogs/nstv/2011/04/ferrofluid-sculptures-reveal-magnetic-fields.html`, 2011. [Online; accessed 22.01.2015].
4. Carter, T., Seah, S. A., Long, B., Drinkwater, B., and Subramanian, S. Ultrahaptics: Multi-point mid-air haptic feedback for touch surfaces. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, ACM (New York, NY, USA, 2013), 505–514.
5. Gupta, S., Morris, D., Patel, S. N., and Tan, D. Airwave: Non-contact haptic feedback using air vortex rings. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '13, ACM (New York, NY, USA, 2013), 419–428.
6. Hachisu, T., and Fukumoto, M. Vacuumtouch: Attractive force feedback interface for haptic interactive surface using air suction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, ACM (New York, NY, USA, 2014), 411–420.
7. Hasegawa, K., and Shinoda, H. Aerial display of vibrotactile sensation with high spatial-temporal resolution using large-aperture airborne ultrasound phased array. In *World Haptics Conference (WHC), 2013* (April 2013), 31–36.
8. Hoshi, T. Compact ultrasound device for noncontact interaction. In *Proceedings of the 9th International Conference on Advances in Computer Entertainment*, ACE'12, Springer-Verlag (Berlin, Heidelberg, 2012), 502–505.
9. Hoshi, T., Takahashi, M., Iwamoto, T., and Shinoda, H. Noncontact tactile display based on radiation pressure of airborne ultrasound. *EEE Trans. Haptics 3*, 3 (#jul# 2010), 155–165.
10. Hoshi, T., Takahashi, M., Iwamoto, T., and Shinoda, H. Noncontact tactile display based on radiation pressure of airborne ultrasound. *Haptics, IEEE Transactions on 3*, 3 (July 2010), 155–165.
11. Hoshi, T., Takahashi, M., Nakatsuma, K., and Shinoda, H. Touchable holography. In *ACM SIGGRAPH 2009 Emerging Technologies*, SIGGRAPH '09, ACM (New York, NY, USA, 2009), 23:1–23:1.
12. Inoue, S., Kobayashi-Kirschvink, K. J., Monnai, Y., Hasegawa, K., Makino, Y., and Shinoda, H. Horn: The hapt-optic reconstruction. In *ACM SIGGRAPH 2014 Emerging Technologies*, SIGGRAPH '14, ACM (New York, NY, USA, 2014), 11:1–11:1.
13. Iwaki, S., Morimasa, H., Noritsugu, T., and Kobayashi, M. Contactless manipulation of an object on a plane surface using multiple air jets. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on* (May 2011), 3257–3262.

14. Iwamoto, T., Tatezono, M., Hoshi, T., and Shinoda, H. Airborne ultrasound tactile display. In *ACM SIGGRAPH 2008 New Tech Demos*, SIGGRAPH '08, ACM (New York, NY, USA, 2008), 1:1–1:1.
15. Karunanayaka, K., Siriwardana, S., Edirisinghe, C., Nakatsu, R., and Gopalakrishnakone, P. Magnetic field based near surface haptic and pointing interface. In *Proceedings of the 15th International Conference on Human-Computer Interaction: Interaction Modalities and Techniques - Volume Part IV*, HCI'13, Springer-Verlag (Berlin, Heidelberg, 2013), 601–609.
16. Khalil, K. S., Mahmoudi, S. R., Abu-dheir, N., and Varanasi, K. K. Active surfaces: Ferrofluid-impregnated surfaces for active manipulation of droplets. *Applied Physics Letters 105*, 4 (2014), –.
17. Lee, J., Post, R., and Ishii, H. Zeron: Mid-air tangible interaction enabled by computer controlled magnetic levitation. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, UIST '11, ACM (New York, NY, USA, 2011), 327–336.
18. Long, B., Seah, S. A., Carter, T., and Subramanian, S. Rendering volumetric haptic shapes in mid-air using ultrasound. *ACM Trans. Graph. 33*, 6 (Nov. 2014), 181:1–181:10.
19. Marshall, M., Carter, T., Alexander, J., and Subramanian, S. Ultra-tangibles: Creating movable tangible objects on interactive tables. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, ACM (New York, NY, USA, 2012), 2185–2188.
20. Monnai, Y., Hasegawa, K., Fujiwara, M., Yoshino, K., Inoue, S., and Shinoda, H. Haptomime: Mid-air haptic interaction with a floating virtual screen. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, ACM (New York, NY, USA, 2014), 663–667.
21. Nedelcu, M. Expanded image spaces. from panoramic image to virtual reality, through cinema. *Close Up: Film and Media Studies*, 44.
22. Obrist, M., Seah, S. A., and Subramanian, S. Talking about tactile experiences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, ACM (New York, NY, USA, 2013), 1659–1668.
23. Ochiai, Y., Hoshi, T., and Rekimoto, J. Pixie dust: Graphics generated by levitated and animated objects in computational acoustic-potential field. *ACM Trans. Graph. 33*, 4 (#jul# 2014), 85:1–85:13.
24. Otsubo, M. Optical imaging apparatus, Sept. 11 2014. US Patent App. 14/352,616.
25. Popper, B. Cyborg America: inside the strange new world of basement body hackers. `http://www.theverge.com/2012/8/8/3177438/cyborg-america-biohackers-grinders-body-hackers`. [Online; accessed 16.12.2014].
26. Rakkolainen, I. How feasible are star wars mid-air displays. In *Proceedings of the 11th International Conference Information Visualization*, IV '07, IEEE Computer Society (Washington, DC, USA, 2007), 935–942.
27. Rasmussen, M. K. Magical realities in interaction design. In *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction*, TEI '13, ACM (New York, NY, USA, 2013), 125–128.
28. Rasmussen, M. K., Grönvall, E., Kinch, S., and Petersen, M. G. "it's alive, it's magic, it's in love with you": Opportunities, challenges and open questions for actuated interfaces. In *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration*, OzCHI '13, ACM (New York, NY, USA, 2013), 63–72.
29. Rey, C. Acoustic levitation and methods for manipulating levitated objects, Aug. 18 1981. US Patent 4,284,403.
30. Sodhi, R., Poupyrev, I., Glisson, M., and Israr, A. Aireal: Interactive tactile experiences in free air. *ACM Trans. Graph. 32*, 4 (#jul# 2013), 134:1–134:10.
31. Suzuki, Y., and Kobayashi, M. Air jet driven force feedback in virtual reality. *Computer Graphics and Applications, IEEE 25*, 1 (Jan 2005), 44–47.
32. Tsalamlal, M., Issartel, P., Ouarti, N., and Ammi, M. Hair: Haptic feedback with a mobile air jet. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on* (May 2014), 2699–2706.
33. Tsalamlal, M., Ouarti, N., and Ammi, M. Psychophysical study of air jet based tactile stimulation. In *World Haptics Conference (WHC), 2013* (April 2013), 639–644.
34. Weiss, M., Schwarz, F., Jakubowski, S., and Borchers, J. Madgets - Actuating Widgets on Interactive Tabletops. `https://hci.rwth-aachen.de/madgets`. Accessed: 1.11.14.
35. Weiss, M., Schwarz, F., Jakubowski, S., and Borchers, J. Madgets: Actuating widgets on interactive tabletops. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, ACM (New York, NY, USA, 2010), 293–302.
36. Weiss, M., Wacharamanotham, C., Voelker, S., and Borchers, J. Fingerflux: Near-surface haptic feedback on tabletops. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, UIST '11, ACM (New York, NY, USA, 2011), 615–620.
37. Wilson, G., Carter, T., Subramanian, S., and Brewster, S. A. Perception of ultrasonic haptic feedback on the hand: Localisation and apparent motion. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, ACM (New York, NY, USA, 2014), 1133–1142.