Volume 9

Katarina Foldenauer, Tobias Plischke, Luca Strauß,
Eike Viehmann, Michael Koch

Schriften zur soziotechnischen Integration

Using Audio for Detecting Covered Users in Front of Public Displays

Forschungsgruppe Kooperationssysteme,
Universität der Bundeswehr München

www.soziotech.org

CSCM
Forschungsgruppe
Kooperationssysteme
München

Universität der Bundeswehr München

ISSN 2194-0274 (Print)
ISSN 2194-0282 (Online)

Bibliographic data according to the German National Library
The German National Library is recording this publication in the German National Bibliography. Detailed bibliographic data is available online:
https://portal.dnb.de

Series: Schriften zur soziotechnischen Integration
Published by:
Michael Koch & Florian Ott

Volume 9: Using Audio for Detecting Covered Users in Front of Public Displays
Contributed by:
Katarina Foldenauer, Tobias Plischke, Luca Strauß, Eike Viehmann, Michael Koch

Referenz / Zitation

Katarina Foldenauer, Tobias Plischke, Luca Strauß, Eike Viehmann, Michael Koch (2022): „Using Audio for Detecting Covered Users in Front of Public Displays". Schriften zur soziotechnischen Integration, Volume 9. München: Forschungsgruppe Kooperationssysteme, Universität der Bundeswehr München.

https://schriften.soziotech.org/band9

ISSN 2194-0274 (Print)
ISSN 2194-0282 (Online)
DOI 10.18726/2022_4

1st edition, September 2022
Cover design: Eva Stuke, Layout: Florian Ott

Forschungsgruppe Kooperationssysteme, Universität der Bundeswehr München
Werner-Heisenberg-Weg, 85577 Neubiberg

E-Mail: schriften@soziotech.org
Internet: https://schriften.soziotech.org

Using Audio for Detecting Covered Users in Front of Public Displays

Katarina Foldenauer, Tobias Plischke, Luca Strauß, Eike Viehmann, Michael Koch

Abstract

Public displays have become more and more ubiquitous in recent years. Both for managing public display networks and for doing research in this field it is an important issue to automatically detect what is happening in front of these screens. This often is done with the help of depth cameras. However, measurement errors can occur due to occlusion or underexposure. In our research we looked into the idea to support the detection of concealed persons using audio information. A series of laboratory experiments showed that audio recordings can be used to detect concealed persons. This can be used in existing installations to minimize occlusion error in visual recordings.

Keywords

Public displays; person detection; audio information; experiment

Table of Contents

Abstract i
Keywords i
Table of Contents ii
Table of Figures iii

1 INTRODUCTION 1

2 RELATED WORK 2

3 EXPERIMENTS 4

3.1 Resolving Occlusion 5
3.2 Walking Speed 6
3.3 Background Noise 7

4 EVALUATION 9

4.1 Frequency Analysis 9
4.2 Evaluation of the Frequency Spectrum 10
4.3 Volume (Arithmetic Mean) 15
4.4 Evaluation of the Arithmetic Mean of the Volume 16

5 DISCUSSION / CONCLUSION 19

BIBLIOGRAPHY 20

LIST OF AUTHORS 21

Table of Figures

Fig. 1: Experiment setup ... 4

Fig. 2: Measurement with walking (three persons) ... 5

Fig. 3: Walking with different speeds ... 6

Fig. 4: Walking with background noise ... 8

Fig. 5: Frequency analysis of arc-shaped movement ... 10

Fig. 6: Frequency analysis of orthogonal motion ... 11

Fig. 7: Frequency analysis parallel movement ... 12

Fig. 8: Frequency analysis with steps in the background ... 13

Fig. 9: Frequency analysis with conversation in the background ... 14

Fig. 10: Frequency analysis with different speeds ... 14

Fig. 11: Comparison Arith. Mean of loudness (dBs) without background noise ... 17

Fig. 12: Comparison Arith. Mean of loudness (dBs) with background noise ... 18

1 Introduction

In recent years public displays have become more and more ubiquitous. Both for managing public display networks and for doing research in this field it is an important issue to automatically detect what is happening in front of these screens – e.g. how many people are watching or passing by.

Visually analyzable recordings with e.g. Microsoft Kinect depth sensors are widely used for the detection of user motion profiles (Gillian et al. 2014). A major problem of visual evaluation with depth sensors is the measurement inaccuracy, which is caused by the obscuration of people, among other things. This effect becomes especially apparent when people move together in groups or offset within the sensor's field of view. Occlusion comes into play with (all) visual sensors, especially with camera-assisted evaluations. The occlusion can be minimized by installing additional (visual) sensors. However, this is not always possible.

This research addresses the question whether audio recordings are an alternative way to solve the "occlusion error" and may even provide additional data. The idea is to record the sounds of activities in front of a public display with a microphone in addition to the visual evaluation.

In the remainder of this paper we will first present related work. Then we will briefly describe the experiments we conducted in which different situations are simulated and recorded with an audio recording device. The different runs aim at reenacting an occluded person in front of a pervasive display. Subsequently, it was examined whether and how, if necessary, the number of persons can be inferred from the audio data. Finally, the recorded data will be evaluated, whereupon a summary will provide both an overview and an outlook.

2 Related Work

A lot of work has been done on detecting people and walking paths in front of public displays using cameras.

A cornerstone for capturing users via video tracking was laid by Yan and Forsyth (2005). With this tracking system, it was possible to automate data analysis instead of having to analyze it through human hands. In their experiment, a public square was analyzed to find out more about the movement patterns and time spent by visitors.

Elhart et al. (2017) developed an open-source tool for tracking in front of public displays. Using a Kinect camera, they were able to detect people in a limited area in front of the display with about 91% accuracy. However, the authors also identified some problems. For example, high-contrast clothing had a negative influence, since the algorithm they used assumed several people instead of one individual. Also, occlusion played a role in groups of people, which in retrospect led to them being perceived as one person.

Sometimes the camera-based solutions have been extended using other sensors. Gillian et al. (2014), for example, combined nearly one hundred sensors to collect data for this purpose. This included RFID readers, as well as depth and RGB cameras. With this information it was possible to create an interactive map showing the position of the people in front of public displays.

In general, little research was done on using audio to derive information about people in front of screens. While there are intelligent systems that, for example, detect voices and perform a specific action in response (Schulte et al. 2001), this data has not been used to determine the number of people actually present or even to track their walking paths.

Meier (1998) uses audiovisual data for the investigation of work and interaction processes and describes hurdles that have to be overcome during the acquisition and evaluation. The work also presents detailed set-up possibilities and deals in detail with various issues like the positioning of the sensors, which can have an influence on the result.

Where and how the microphones are placed depends on their nature and the environment to be recorded. It also depends on the material capacities and can affect the output accordingly. The general operation and placement of these same microphones is described in Elsea (1996).

Auditory stimuli can be a part of public displays to attract attention. Kukka et al. (2016) try to reduce the so-called display blindness by following three different audiovisual approaches. However, so far, these sounds have always been equally loud and sometimes elicited unpleasant reactions. In the same breath, the authors suggested integrating a microphone to detect the volume level of the environment and adjust the volume accordingly. Adapted to this practical work, the microphone could probably also be used to trigger these sounds, if it is possible to assign movements through a microphone.

On a smaller scale, microphones have also been used to detect a patient's cough, for example. Here, too, two microphones and selected algorithms were used to evaluate the audio data in order to be able to sound an alarm for the patient in the event of a cough. Accuracies of up to 95% were achieved with the methods presented here, which was an advance over previous methods (Drugman et al. 2020).

In Hespanhol (2016), microphones were used to perceive the environment and then use this information for audio-visual feedback in the form of projections on the water. Since this project was carried out at a harbor took place, the work also addresses issues such as background noise or the optimal setup of the whole, as this was necessary for the projections to succeed. However, once again, the focus is on interactivity.

When it comes to advertisement, several techniques have been used to either attract attention or even elicit an interaction with the display. In Eken (2017) and Sahibzada et al. (2017) many different sensors, such as ultrasonic sensors, cameras or mobile phones, are being used to detect people, followed up by a certain animation or the possibility to interact with the device. In addition to that, speech recognition and voice commands can be used to improve the interactivity and have been compared to other sensors in Tafreshi et al. (2018).

3 Experiments

To find if it is possible to detect visual occlusion of users by using audio information, a series of laboratory experiments that represent characteristic situations was performed. Audio information was recorded and verified with the known baseline truth. The scenarios for the experiments have been created based on the following criteria:

- Number of persons (1 person - 2 persons - 3 persons) –
  Can occlusion be detected?
- Walking speed (Walking at normal speed - Brisk walking) –
  Can extended statements about motion profile be made?
- Background noise – for more realistic results

Fig. 1: Experiment setup

All experiments were conducted in a 40 square meter room and recorded with a Tascam 2-channel DR-100 MK II audio recorder. Two Sennheiser directional microphones are used for capturing the audio, one each to the left and right of the theoretical “screen". Thus, the focus of the recording is directly in front of the screen. The distance of the microphones follows from the fact that people interacting with the public screen are usually at a distance of one meter from the screen. The microphones were mounted on tripods at a height of 10 cm to capture step sound. These microphones have directional characteristics, i.e. they pick up mainly frontal sound, while sound from other directions is perceived as attenuated. This means that, with the appropriate orientation, certain signals may be better filtered out despite surrounding noise. The audio information was saved in WAV format with a sampling rate of 48 kHz. The sampling depth of the individual sample points was 24 bits.

The experiment was performed in a closed room. The two microphones were attached to the left and right of a fictitious public display (whiteboard) on the floor in a small tripod, so that the screen is placed centrally above it. Every experiment was performed/recorded five times.

A picture of the original setup of the experiment is shown in Figure 1.

3.1 Resolving Occlusion

In the first series of experiments, one, two and finally three persons simultaneously walked along different paths in front of the screen. Figure 2 shows the setup with three persons. The audio recordings of the experiment with one person were used as a basis for later comparison.

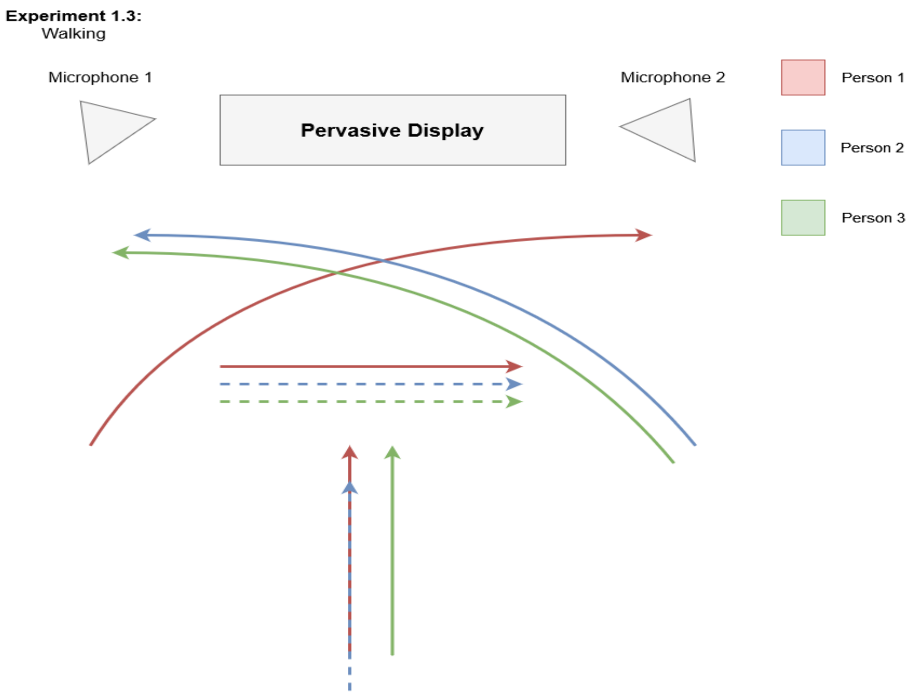

Fig. 2: Measurement with walking (three persons)

At the beginning, the two test persons walked in an arc starting from the left and right towards the screen. Here one of the persons was hidden only for a short moment when crossing the routes. However, due to the different starting points, this can also be recognized by video recordings. Though, the audio track should still be helpful to recognize differences to where a different number of people are moving in front of the screen. Subsequently, the persons walked straight ahead towards the screen one after the other. In the process, one person was obscured by a potential video recording. When walking parallel to the screen, one of the two people was obscured, since in our case they walk side by side at the same speed. The dashed routes in Figure 2 represent the occlusion of this person.

3.2 Walking Speed

In the second set of experiments, the subjects walked past the screen at different speeds. The test person first walked from the right in an arc towards the screen. In the next run, the person walked the same distance, but faster than before. In this experiment, it did not matter which path the person took. The important thing here was that the path is the same at both speeds. In this way, the audio tracks can be optimally compared with each other and a clear result can be expected.

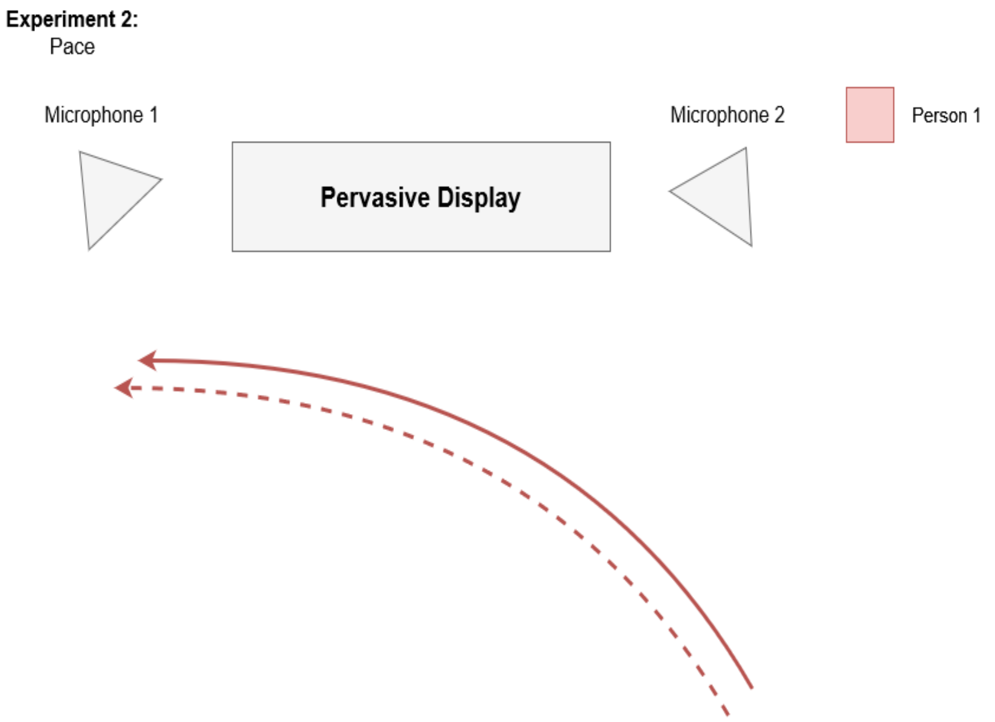

Fig. 3: Walking with different speeds

3.3 Background Noise

In the first series of experiments we have already recorded different scenarios with different numbers of test persons. Thereby there were no intentional noises in the background so far. In order to make the test environment more realistic and to be able to avoid possible errors in the recognition of persons with audio recordings afterwards, we added background noise in this series of tests. We introduced two different types of sounds in the background: First, conversations in the background and then a person moving in the background.

In the first test run we recorded one person with background noise. This was later used for direct comparison with multiple-people scenarios. Similar to the previous series of experiments, a test person walked straight past the screen parallel to it. While the subject walked past the screen, two people were talking in the background. Two people walked past the screen parallel to each other at the same speed. As in the previous experiment, there was a background noise in the form of a conversation. In this experiment, there was a visual occlusion of a person. The recordings were to be used to detect whether the occluded person is recognized by the audio track despite the conversation in the background.

In another experimental run, we recorded the same scenarios as in Figure 2, but this time with a different noise backdrop. As it is often the case in reality, there were other moving people in the background in this experiment. Here, too, we let one and then two persons walk past the screen parallel to the screen. In this way, we wanted to investigate whether the number of people in front of the screen can be detected with the help of audio recordings despite moving people in the background.

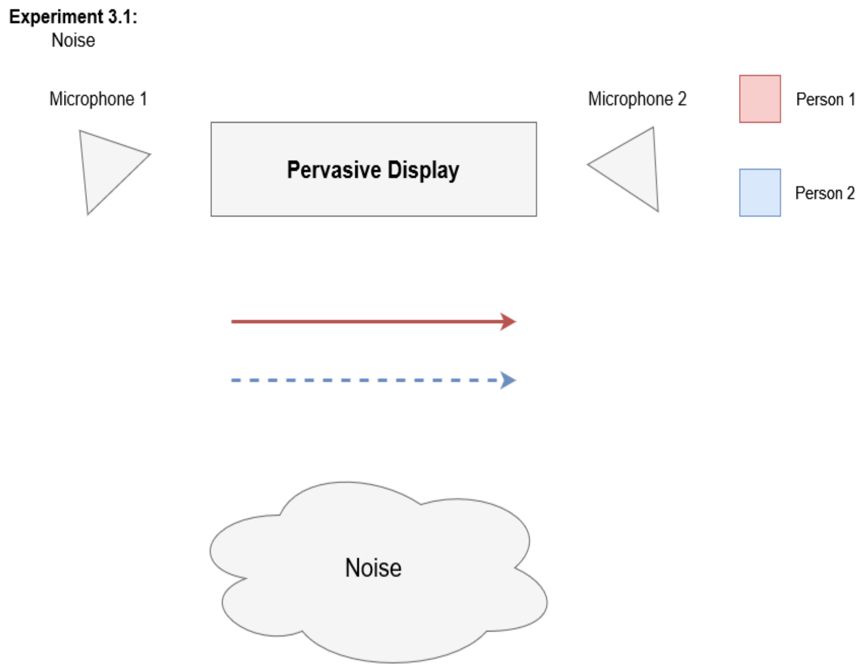

Fig. 4: Walking with background noise

4 Evaluation

In the following, the evaluation of the experiment presented in the previous chapter is described. The evaluation was done by considering two different aspects. First, an attempt was made to make target-oriented statements with the help of a frequency analysis. Then, the volume of the audio files was considered. The goal of this chapter is to find out if and which statements can be made based on the above-mentioned aspects.

4.1 Frequency Analysis

In the first step of the evaluation of the experiment we compared our audio recordings on the basis of a frequency analysis to derive first clues with regard to our research question. The frequency spectrum provides information about the composition of the frequencies of an audio recording. From a physical point of view, an audio signal is a superposition of many individual frequencies, which are classified into low pass (20-40 Hz), mid bass (40-100 Hz), upper bass (100-150Hz), lower mid (150-400 Hz), middle mid (400 Hz - 1 kHz), upper mid (1-2 kHz), lower high (2-3.5 kHz), middle high (3.5-6 kHz), upper high (6-10 kHz) and super high (10-20 kHz).

Altogether, a tone consists of a superposition of any number of frequencies from the frequency range 20-20000 Hz, which characterize a tone in its entirety. With the help of the frequency spectrum of the audio signal, the characteristics of the recorded soundscape (composition of the frequency response) can be mapped. In particular, when analyzing the frequency spectrum, we expect the frequency response of the audio recording to become "fuller" as the number of people increases, i.e., the level values in the primarily lower range (bass and midrange) of the frequency spectrum increase.

The frequency spectrum can be used to make an initial qualitative statement about the (maximum) level values of individual frequencies in the frequency spectrum. Insofar as in the comparison of the frequency spectra of the test series in the considered frequency range the levels increase with the number of running persons, a first (weak) statement can be made about the assumption "loudness correlates with number of running persons". However, since the frequency spectrum does not allow a quantitative statement about all levels of the audio signal, but only always represents the qualitatively largest value of individual frequencies, the frequency spectrum is not suitable to derive a statement about the (average) volume of the audio track. For this reason, in the next section we want to calculate the arithmetic mean of the loudness of the audio recordings and then compare them in the context of the research question.

4.2 Evaluation of the Frequency Spectrum

For the following evaluations the program "Audacity" was used and the results from it were presented graphically with the help of Excel. For an optimal frequency analysis, the five individual tracks of each scenario were first averaged. With the frequency analysis tool of Audacity the audio tracks were analyzed and could be exported into a text file. The resulting graphs are explained in more detail below. A frequency analysis provides information about the proportion in which certain frequencies are represented in a signal.

As in the experimental procedure, we first start with the arc-shaped paths of the test subjects (see Figure 2). Since there are no covered subjects here, we will first consider the difference in the number of subjects based on the frequency spectrum.

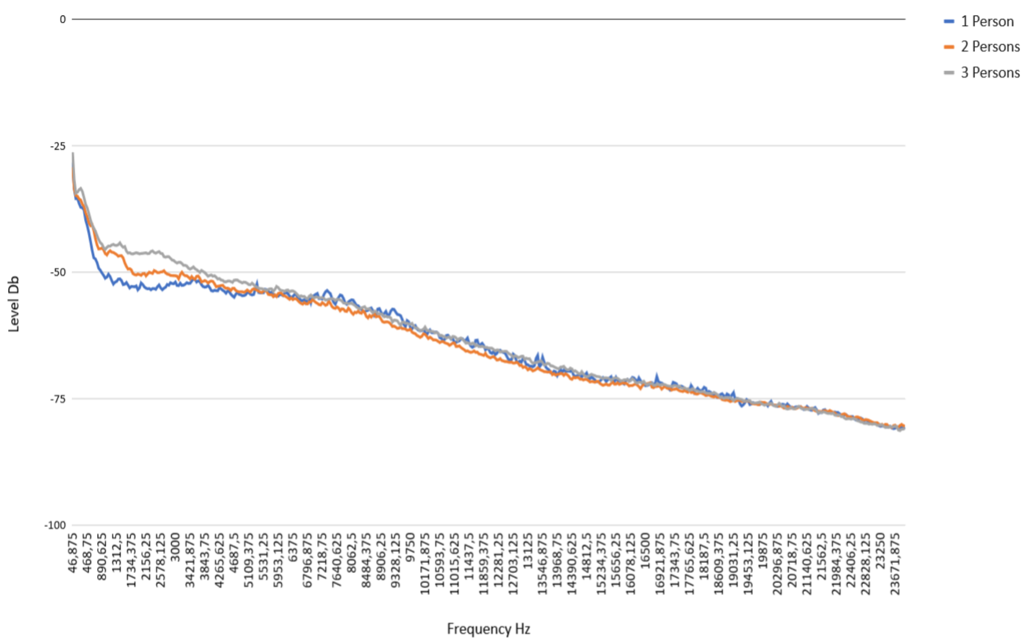

Fig. 5: Frequency analysis of arc-shaped movement

The frequency diagram in Figure 5 shows the frequency curves of the arc-shaped path. The number of subjects is indicated by the colors blue, orange and gray and are thus in direct comparison. In the graph it becomes clear that for three persons especially the frequencies in the range of about 700 Hz to 5000 Hz occur more frequently than for two persons. The same applies to two persons compared to one person. So, we can already see the first differences in the frequency analysis here. To answer our research question concretely, we now consider the paths of the test subjects in which occluded persons occur. This includes walking orthogonal to the screen and walking parallel to the screen. Figure 6 shows the frequency spectrum of the audio recordings of one, two, and three subjects walking orthogonally to the screen, respectively.

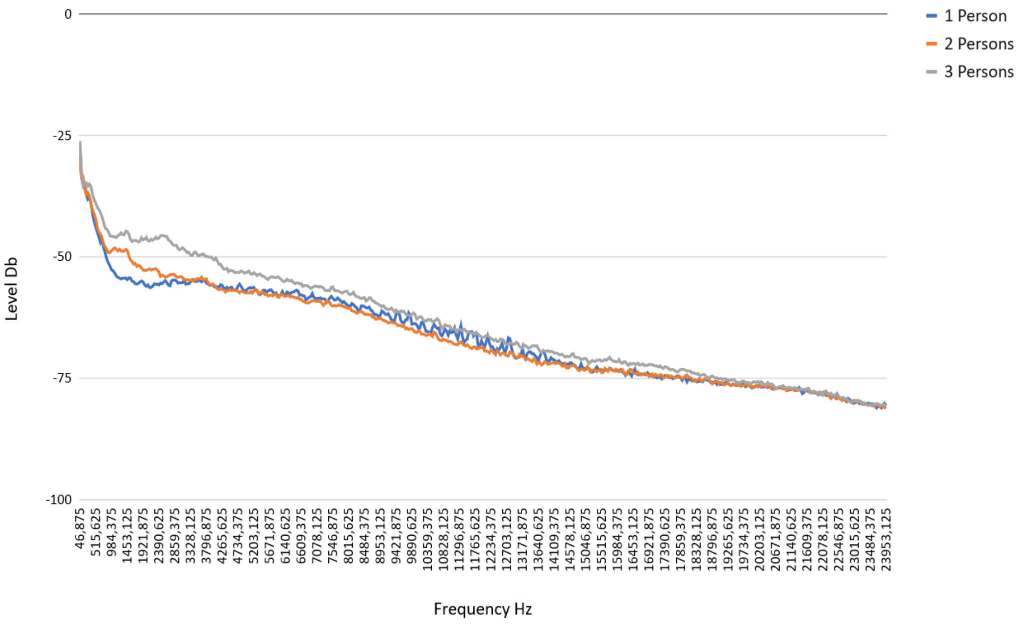

Fig. 6: Frequency analysis of orthogonal motion

Similar to the previous scenario, the frequencies from 900 Hz to 3000 Hz occur more frequently for three and two people. Also, overall, all frequencies occur consistently more often with three people. However, in the middle of the graph, it can be seen that some frequencies occur more frequently with one person than with two people.

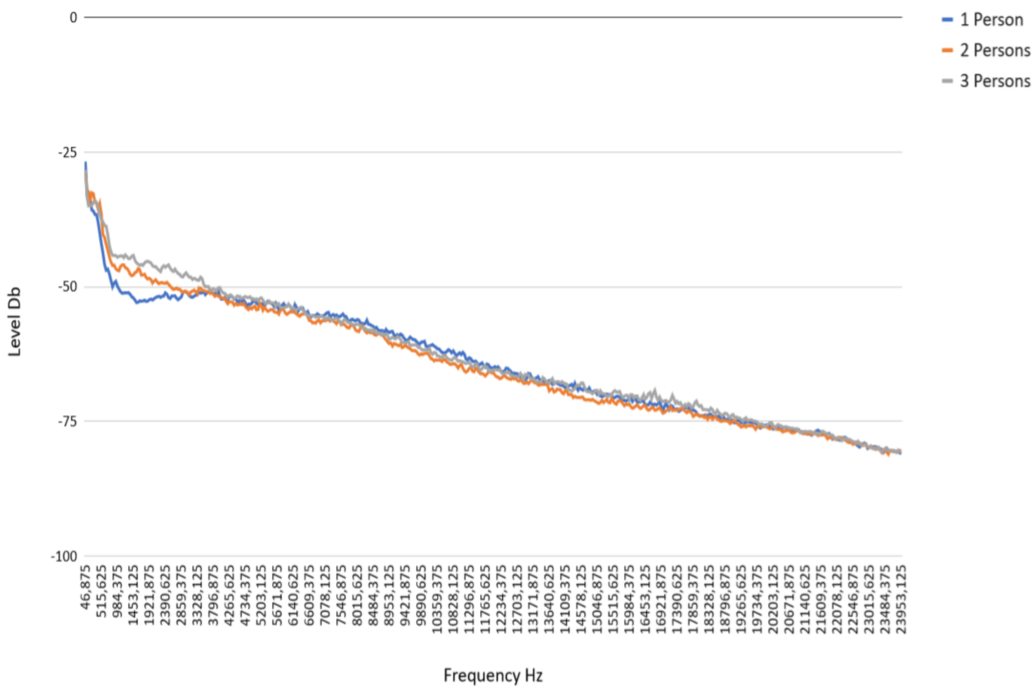

Fig. 7: Frequency analysis parallel movement

Even when moving parallel to the screen, the patterns repeat. In the front area, the graph of three persons shows the most frequencies, followed in descending order by the graphs of two and one person. In direct comparison to the first test series, we now consider the frequency analyses with noise backgrounds. Here we want to observe whether the background noise has an influence on the recognition of the number of persons. For this purpose we compare one and two test persons as they pass the screen in parallel.

Fig. 8: Frequency analysis with steps in the background

Figure 8 shows the described scenario with walking people in the background. The graphs show that almost all frequencies occur more frequently with two people in front of the screen than with only one person, the same can be seen with a conversation in the background (see Figure 9). Thus, there is no distortion of the number of persons visible in the frequency spectrum due to moving persons in the background.

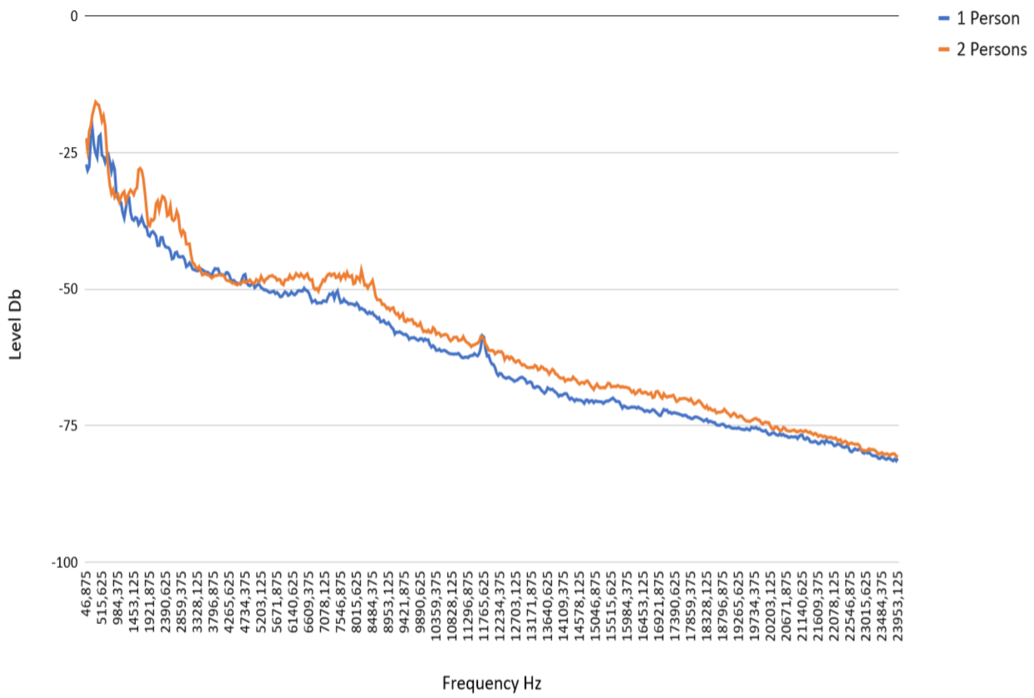

Fig. 9: Frequency analysis with conversation in the background

In the second experiment, a test person was recorded walking at two different speeds. Figure 10 represents the frequency analysis of the two audio recordings.

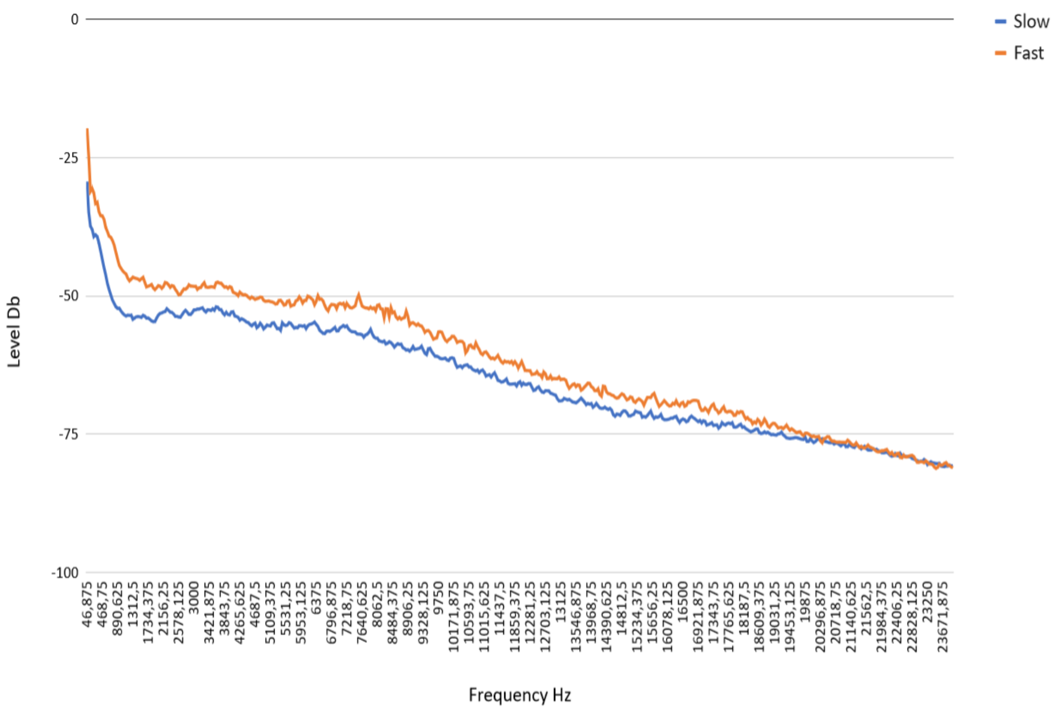

Fig. 10: Frequency analysis with different speeds

Here, a clear difference can be seen in the two differently colored curves. At a higher speed, all frequencies clearly occur more frequently than during normal walking.

Since this was only a first graphical analysis of the frequency spectrum, we observed the volume of the individual audio tracks, more precisely their average values. In the following subchapter this method will be discussed in more detail.

4.3 Volume (Arithmetic Mean)

The Tascam audio recorder exports the recorded 2-channel recordings into the WAV container format. Here, the left channel corresponds to the audio recording of the left Sennheiser microphone and the right channel corresponds to the audio recording of the right Sennheiser microphone. Since in the context of the research question and in this subsection, we are particularly examining the loudness for robust statements regarding the number of moving people, in the following we consider the arithmetic mean of the loudness of characteristic and comparable experimental performances. In particular, we want to investigate here whether the number of moving persons can be distinguished based on the arithmetic mean of the loudness of the recorded stereo audio tracks.

To calculate the arithmetic mean of a stereo audio recording, we first calculate the average of the loudness of two temporally corresponding sample points from the left and right audio tracks. Then we calculate the arithmetic mean of all average values from the left and right channels. The Tascam audio recorder has a maximum sample rate of 64 k/s. For this reason, 64 k data points each from the right and left channels are recorded per second. Since the Tascam audio recorder uses the logarithmic unit decibels relative to full scale (dBFS) with absolute scale, the value 0 dBFS represents the highest-level value (volume), whereas all lower volumes are represented with negative dBFS.

Assuming a stereo audio track was recorded for only one second, the arithmetic mean of this stereo audio track is calculated by first averaging the left and right channels (adding the two temporally corresponding sample points from the left and right channels and multiplying by 1/2). The result is then an audio track with one channel and 64 k sample points, where each sample point now represents the average volume from the left and right audio channels. To calculate the arithmetic mean of the total volume of this original stereo audio track, all individual 64 k sample points are now added together and divided by the number of sample points (64 k). The result is a dBFS value that represents the arithmetic mean of the volume of the stereo audio recording.

We would like to point out again at this point that audio recordings are basically always very individually dependent on the hardware, the setting of the microphones and the environment in which they were recorded. For this reason, the recordings made and the laboratory tests can only be compared in the context of our test series. In particular, we made sure that all experiments were recorded with the same microphone setup and settings and the same basic spatial conditions (microphones always in the same place, no acoustic changes such as windows/doors open/closed or furniture moved).

4.4 Evaluation of the Arithmetic Mean of the Volume

In order to evaluate the loudness of the sample points, the samples with the decibel measurement scale had to be extracted from Audacity using the sample data export. Since not all recordings were free of unwanted noise, such as the announcement of the respective experiment, they had to be cropped accordingly beforehand. To ensure that the samples remained comparable, care was taken to ensure that the cutting started between the announcement and the run, as there was always a short pause there.

For the export, the number of samples still had to be determined. Here, 330,000 points were used to ensure that all data points would be taken into account even in somewhat longer trials. Afterwards, all decibel values of the sample points for both tracks of the microphones were saved in a .txt file.

For the evaluation, the average value of all runs of an experiment remains, which is compared with the value of other experiments in the following.

For the tests without noise background, values between -53 and -43 resulted and are shown in Table 1. The results of the table are shown graphically in Figure 11.

	1 person	2 persons	3 persons
Orthogonal	-52,356967	-48,84051662	-44,16342451
Parallel	-45,51962658	-46,95659898	-44,89983429
Curve	-46,66618623	-45,21823141	-43,95705714

Table 1: Comparison without background noise

When these values are plotted graphically, it is clear that the averages of the loudness differ for the respective experiments. Especially in the experiment with the orthogonal path in Figure 11, clear volume jumps can be seen. Here, it is thus possible to distinguish between the number of persons at least from the point of view of the decibel values with a constant set-up.

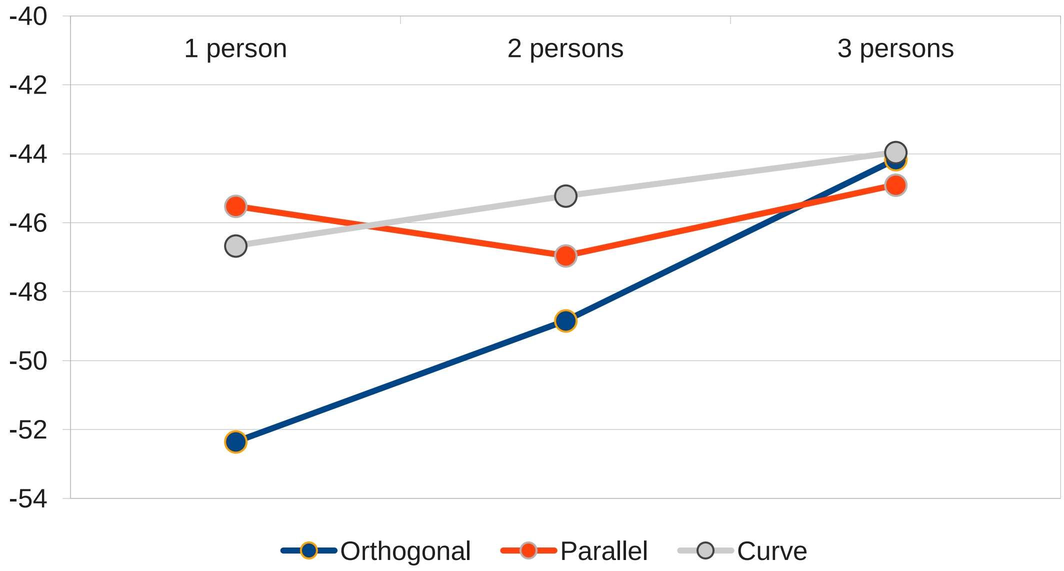

Fig. 11: Comparison Arith. Mean of loudness (dBs) without background noise

While the orthogonal test still shows an amplitude of about 8 decibels, the data points in the curvilinear test differ by only about 2.5 decibels. Although the increase in loudness with the number of people is still noticeable, it is smaller than in the former.

The test with parallel runs, marked in orange in Figure 10, falls out of the pattern:

According to the recorded samples, here on average two people were quieter than one or three people. In general, the data points are closest together with an amplitude of about 2 decibels. The reason for this irregularity could be, on the one hand, a different running behavior during the recording, as sometimes large differences can be seen here in the 5 individual runs. The complete table is attached in the appendix. On the other hand, it could also be due to the running path itself, since here at all times the persons run one behind the other across the microphones, unlike in the other experiments.

The comparison of the loudness with a background noise resulted in values between -43 and -31 decibels and can be seen in Table 2.

	1 person	2 persons
Parallel (steps in the background)	-42,70212608	-35,09892124
Parallel (conversations in the background)	-39,18277126	-31,03002511

Table 2: Comparison with background noise

In the corresponding graph in Figure 11 you can clearly see that the loudness increases as the number of people increases for both experiments.

The differences between the data points here range from 3.5 to 4 decibels, and the trials were generally louder than those without the background noise.

Furthermore, it can be seen that the music in the background was louder than the footsteps in the background, contributing to an overall louder average.

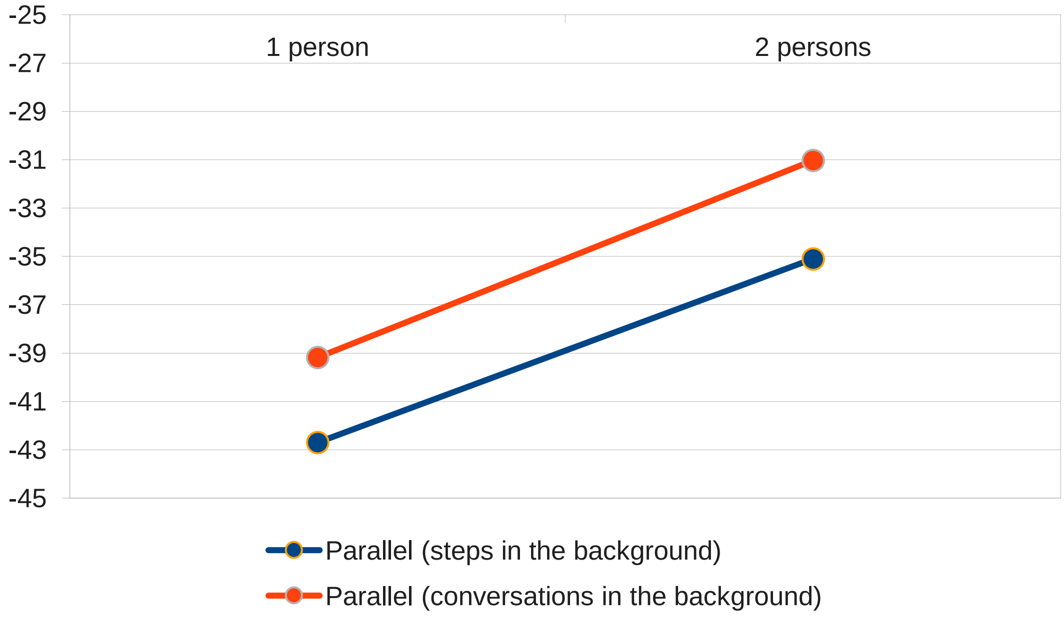

Fig. 12: Comparison Arith. Mean of loudness (dBs) with background noise

5 Discussion / Conclusion

The execution of the tests and their evaluated measurement data showed that they support the detection of concealed persons. With the direct comparison of different numbers of persons and the same paths in front of the screen, conclusions could be drawn about the number of persons. With an increasing number of persons, a significantly higher number of occurring frequencies was detected. In addition, we were also able to prove that the arithmetic mean of the volume correlates with the number of persons. With these results, we can thus prove that audio recordings of motion profiles enable the detection of concealed persons. This can be used in existing installations (such as pervasive displays) to minimize occlusion error in visual recordings.

Despite this measurement methodology being tested for the first time, it was possible to measure targetable results, which can be used as a basis for further investigations in the future. Since clear differences could be seen in the results of tests with different numbers of persons, tests with varying background noise as well as tests outside of buildings can be carried out based on these results. When evaluating the data, it should be noted that the five consecutively recorded audio files of an experiment (e.g. orthogonal) sometimes show fluctuations. Therefore, it is recommended to collect larger data sets for more accurate measurement results.

Since we could only access a number of three test persons in our experiment, it remains to be investigated how the measurement data behave in experiments with more participants. In addition, the influence of the number and position of the microphones on the experimental results must be considered.

Bibliography

DRUGMAN, T.; URBAIN, J.; BAUWENS, N.; CHESSINI, R.; AUBRIOT, A.-S.; LEBECQUE, P. AND DUTOIT, T. (2020) "AUDIO AND CONTACT MICROPHONES FOR COUGH DETECTION".

ELSEA, P. (1996). HTTP://ARTSITES.UCSC.EDU/EMS/MUSIC/TECH_BACKGROUND/TE-20/TECES_20.HTML.

EKEN, B. (2017) "INTERACTIVE DIGITAL DISPLAYS FOR OUTDOOR ADVERTISING. NEW TRENDS AND ISSUES PROCEEDINGS ON HUMANITIES AND SOCIAL SCIENCES". 11, 4 (2017), 85–92. DOI:HTTPS://DOI.ORG/10.18844/PROSOC.V4I11.2853.

ELHART, I.; MIKUSZ, M.; MORA, C. G. AND DAVIES, N. (2017) "AUDIENCE MONITOR: AN OPEN SOURCE TOOL FOR TRACKING AUDIENCE MOBILITY IN FRONT OF PERVASIVE DISPLAYS," IN PERDIS '17: PROCEEDINGS OF THE 6TH ACM INTERNATIONAL SYMPOSIUM ON PERVASIVE DISPLAYS, 2017, S. 1–8.

GILLIAN, N.; PFENNINGER, S.; RUSSELL S. AND PARADISO, J. (2014) "GESTURES EVERYWHERE: A MULTIMODAL SENSOR FUSION AND ANALYSIS FRAMEWORK FOR PERVASIVE DISPLAYS," PERDIS 2014 - PROCEEDINGS: 3RD ACM INTERNATIONAL SYMPOSIUM ON PERVASIVE DISPLAYS 2014, JUNI 2014.

HESPANHOL, L. (2016) "INTERACTING WITH LAUGHTER: A CASE STUDY ON AUDIO-BASED INTERACTI- VITY OF PUBLIC PROJECTIONS," IN PROCEEDINGS OF THE 28TH AUSTRALIAN CONFERENCE ON COMPUTER-HUMAN INTERACTION, SER. OZCHI '16, NEW YORK, NY, USA: ASSOCIATION FOR COMPUTING MACHINERY, 2016, 205–214.

KUKKA, H.; GONCALVES, J.; WANG, K.; PUOLAMAA, T.; LOUIS, J.; MAZOUZI, M. AND ROA BARCO, L. (2016) "UTILIZING AUDIO CUES TO RAISE AWARENESS AND ENTICE INTERACTION ON PUBLIC DISPLAYS," JUNE 2016.

MEIER, C. (1998) "ZUR UNTERSUCHUNG VON ARBEITS- UND INTERAKTIONSPROZESSEN ANHAND VON VIDEOAUFZEICHNUNGEN". ARBEIT. 7:3, 257–275.

SAHIBZADA, H. ET AL. 2017. DESIGNING INTERACTIVE ADVERTISEMENTS FOR PUBLIC DISPLAYS. CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS - PROCEEDINGS (2017), 1518–1529.

SCHULTE, O. A.; FRIEBEL, M. AND KLOTZEK, C. (2001) "AUFZEICHNUNG TECHNISCH VERMITTELTER KOMMUNIKATION – DAS BEISPIEL VIDEOKONFERENZ," GESPRÄCHSFORSCHUNG - ONLINE- ZEITSCHRIFT ZUR VERBALEN INTERAKTION, NR. 2, 222–242, 2001.

YAN, W. AND FORSYTH, D. A. (2005) "LEARNING THE BEHAVIOR OF USERS IN A PUBLIC SPACE THROUGH VIDEO TRACKING," IN 2005 SEVENTH IEEE WORKSHOPS ON APPLICATIONS OF COMPUTER VISION (WACV/MOTION'05) - VOLUME 1, BD. 1, 2005, S. 370–377.

List of Authors

Katarina Foldenauer, Tobias Plischke, Luca Strauß, Eike Viehmann

katarina.foldenauer@unibw.de, tobias.plischke@unibw.de
lcstrss@posteo.de, eikeviehmann@outlook.de

Katharina Foldenauer, Tobias Plischke, Luca Strauß, Eike Viehmann have finished their Bachelor degree in Informatics at Universität der Bundeswehr München and are currently finishing their Master degree in Informatics at the same university.

Prof. Dr. Michael Koch

michael.koch@unibw.de
www.unibw.de/inf2/personen/professoren/univ-prof-dr-michael-koch

Prof. Dr. Michael Koch studied computer science at the TU Munich and received his doctorate in the subject. After an industrial stay at the Xerox Research Centre Europe and subsequent habilitation in computer science again at the TU Munich, he now teaches at the University of the Federal Armed Forces Munich where he holds the professorship for Human-Computer Interaction.

Using Audio for Detecting Covered Users in Front of Public Displays

Public displays have become more and more ubiquitous in recent years. Both for managing public display networks and for doing research in this field it is an important issue to automatically detect what is happening in front of these screens. This often is done with the help of depth cameras. However, measurement errors can occur due to occlusion or underexposure. In our research we looked into the idea to support the detection of concealed persons using audio information. A series of laboratory experiments showed that audio recordings can be used to detect concealed persons. This can be used in existing installations to minimize occlusion error in visual recordings.