ILLUMISENSE: CONTEXT SENSITIVE ILLUMINATION

Marius H. Hennecke *Christian Kleine-Cosack* *Gernot A. Fink*

Intelligent Systems Group, Robotics Research Institute, TU Dortmund University, Germany
{*marius.hennecke, christian.kleine-cosack, gernot.fink*}@*tu-dortmund.de*

ABSTRACT

In order to allow smart technologies to effectively support the user in daily application there is a need for a paradigm shift from direct and explicit to so-called ambient human computer interaction. In this paper we present such a system which implicitly interacts with multiple users in a smart conference room. Based on audio source localization, this system automatically controls the lighting conditions and directs the attention of the audience to the active speakers.

Index Terms— ambient interaction, smart environment, illumination control, acoustic source localization

1. INTRODUCTION

Over the recent years, the role, variety and application of embedded technologies dramatically changed. Allowing, e.g., for intelligent controlling of automation processes in the domestic domain, as well as for interconnecting information over social networks these technologies already invade our daily lives. In such scenarios, the user interaction is based on a multitude of specific user interfaces, utilizing mobile devices and gesture like interaction concepts. However, with more and more applications emerging into our lives, the multitude of direct controlling interfaces will overburden the user and demand for a paradigm shift from direct to ambient user interaction.

In the context of ambient intelligence, different smart home projects[1,2] already show the immense spectrum of technologies and possible applications available [1]. As of today, these projects present solutions for the given, specific tasks on a technical level. But missing the integration of the user and the user's behavior, missing a user oriented interaction, these technology driven projects will conceptually fail for real world applications.

The concept of ambient interaction promises to overcome the aforementioned limitations. The context and the user activities implicitly determine the system's behavior, without the demand for any direct user control and without the need for any visible interface. As a result, the user is supported during daily tasks without the burden of an additional, specific control interfaces.

[1]Living Tomorrow `http://www.livingtomorrow.com`
[2]InHaus `http://www.inhaus-zentrum.de/`

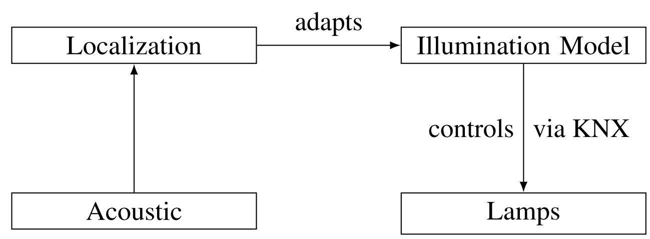

Fig. 1. IllumiSense system architecture

In this paper we present a simple yet illustrative demonstration of ambient interaction. Utilizing acoustic sensors in a smart environment, we use the obtained information to automatically illuminate the scene accordingly. In a smart conference room, this system continuously reacts to the conversational setting, highlights the active speakers and directs the attention of the audience. The goal of this demonstration is to show how users intuitively and implicitly interact with the smart environment. The simplified architecture of the presented system is depicted in Fig.1. The system is real-time capable and used in the field.

2. REALIZATION

The IllumiSense system is realized in a smart conference room located at the Robotics Research Institute at TU Dortmund University. The room serves as a research and experimental platform for intelligent systems design. It contains a variety of different sensor types from which the microphones are used for localizing speakers. In total 16 omni-directional Behringer ECM 8000 microphones are employed for the acoustic source localization (ASL) task. The sensors are connected to two linked Delta 1010 eight-channel sound cards via off-the-shelf pre-amplifier boards. The necessary acoustic signal processing runs on commodity hardware. Additionally, the room is equipped with a multitude of sensors and actors to, e.g., monitor movements and air quality, as well as to control dimmable lamps and roller shutters. Based on the KNX[3] technology, these devices can be accessed via a standardized interface and

[3]KNX `http://www.knx.org`

are connected to the underlying middle-ware framework of the smart environment.

ASL is achieved via the Steered Response Power (SRP) approach [2] which we will review shortly. The main idea is to maximize the output of an acoustic beamformer steered to all relevant positions in the environment. It can be shown that the SRP of a Delay-and-Sum beamformer can be expressed as a sum of cross-correlations $R_{ij}(\tau)$ between channels i and j. In practice the generalized cross-correlation (GCC) with the so-called phase-transformation (PHAT) is commonly employed, which is known to increase the robustness of the cross-correlation estimates against room reverberation. The SRP-PHAT $P(q) = \sum_{(i,j)\in\mathcal{P}} R_{ij}(\tau_{ij}(q))$ for a spatial position q using a total of M sensors is the sum of GCC-PHAT pairs $(i,j) \in \mathcal{P} \subseteq \{1, 2, \ldots, M\}^2$ evaluated for the pairs' time-difference of arrivals $\tau_{ij}(q) = c^{-1}(\|q - p_i\| - \|q - p_j\|)$. The respective microphone positions are denoted by $p_{(\cdot)}$, $\|\cdot\|$ is an Euclidean distance and $c \approx 343\,\mathrm{m\,s^{-1}}$ is the speed of sound. The position $\hat{q} = \arg\max_q P(q)$ for which the SRP function reaches its maximum is an estimate for the dominant acoustic source. The ASL system uses the 16 sensors grouped into two ceiling mounted circular microphone arrays with $0.2\,\mathrm{m}$ diameter each. A three-dimensional grid search of $P(q)$ gives one localization event every $0.15\,\mathrm{s}$ if the acoustic energy is above a pre-defined activation threshold. Figure 2 shows exemplarily the result of this grid search for one time instance.

The ASL results are sent to the OSGi middle-ware[4] which coordinates the communication between different functional blocks of the smart conference room. The presented IllumiSense system is one of several pluggable modules in this dynamic software framework. The number of events per position are accumulated in a coarse grid and form an activity region based illumination model. The OSGI middle-ware updates the actual illumination in the room every $0.4\,\mathrm{s}$ according to the model. The individual distances of lamps to activity regions are mapped to dim values and finally the lamp control commands are send via KNX. On every update all accumulated events are decreased by a fixed amount, leading to a short-time history of activity regions. The accumulated events per grid position are constrained by a maximum number of events, which in combination with the linear decrease per update leads to a time-to-live (TTL) model of activity regions. In the actual implementation the parameters are chosen such that a region is illuminated for a maximum of $30\,\mathrm{s}$ after inactivity. However, the TTL depends on the time of activity beforehand, e.g., a person making a short remark will be illuminated for a short time due to the accumulative nature of the model.

The IllumiSense system described allows for a speaker-length dependent illumination. A demonstration video showing the real-time capabilities and the illumination of multiple speakers is provided on the authors website[5].

[4]OSGi `http://www.osgi.org`

[5]IllumiSense `http://www.irf.tu-dortmund.de/cms/en/IS/Research/SE`

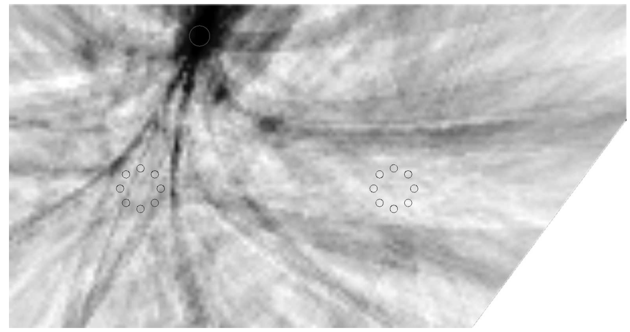

Fig. 2. Visualization of the ASL in the non-rectangular shaped conference room. Microphones are marked by small circles. The source is situated in the middle of the dark area.

3. FUTURE PERSPECTIVES

The system presented here is able to control the illumination of a conference room based on acoustic activity regions. An extension of the localization part of the system to a broader formulation of activity via combining acoustic and visual modalities would allow for an enhanced user experience, e.g., a multi-camera multi-microphone attention [3]. Another possible extension of the IllumiSense system is an automatic context classification [4] based on the ASL results. Integrating such a classification mechanism makes it possible to change the illumination model depending on the automatically inferred context. A prominent example is a presentation scenario where just the presenter should be illuminated and not noisy attendees. Consequently, the illumination model for a presentation or other automatically inferred situations could be chosen accordingly.

4. REFERENCES

[1] J. C. Augusto, “Past, present and future of ambient intelligence and smart environments,” in *Agents and Artificial Intelligence*, ser. CCIS. Springer, 2010, vol. 67, pp. 3–15.

[2] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, “Robust localization in reverberant rooms,” in *Microphone Arrays*. Springer, 2001, ch. 8, pp. 157–180.

[3] B. Schauerte, J. Richarz, T. Plötz, C. Thurau, and G. A. Fink, “Multi-modal and multi-camera attention in smart environments,” in *Proc. Int. Conf. Multimodal Interfaces and Workshop on Machine Learning for Multi-modal Interaction (ICMI-MLMI)*, 2009, pp. 261–268.

[4] C. Kleine-Cosack, M. H. Hennecke, S. Vajda, and G. A. Fink, “Exploiting acoustic source localization for context classification in smart environments,” in *Ambient Intelligence, LNCS*. Springer, 2010, vol. 6439, pp. 157–166.