Identification of Relevant Multimodal Cues to enhance Context-Aware Hearing Instruments

Bernd Tessendorf, Andreas Kettner, Daniel Roggen, Thomas Stiefmeier
Gerhard Tröster
Wearable Computing Lab., ETH Zurich
8092 Zurich, Switzerland
{lastname}@ife.ee.ethz.ch

Peter Derleth, Manuela Feilner
Phonak AG
Laubisrütistrasse 28
8712 Stäfa, Switzerland
{firstname.lastname}@phonak.com

ABSTRACT

Today's state-of-the-art hearing instruments (HIs) adapt the sound processing only according to the user's acoustic surrounding. Acoustic ambiguities limit the set of daily life situations where HIs can support the user adequately. State-of-the-art HIs feature body area networking capabilities. Thus, body-worn sensors could be used to recognize complex user contexts and enhance next-generation HIs. In this work, we identify in a rich real-world data set the mapping between the context of the user –which can be recognized from body-worn sensors– and the user's current hearing wish. This is the foundation for the implementation of recognition systems for the specific cues in next generation HIs based on on-body sensor data. We discuss how the identified mapping allows selecting a-priori distributions for hearing wishes and HI parameters like the switching sensitivity. We conclude deducing the sensory requirements to realize next generation of networked HIs.

Categories and Subject Descriptors

C.3 [**Special-Purpose And Application-Based Systems**]: Signal processing systems

General Terms

Hearing Instrument Body Area Network, Multimodal Sensing, Real-Life Study

1. INTRODUCTION

In collaboration with a hearing instrument (HI) manufacturer we investigate how enhanced contextual awareness can lead to a next generation of HIs with a more effective automatic selection of the hearing program[1] according to the location of the user, their posture, modes of locomotion, and high-level activities. There is a strong need for automatic hearing program selection as manual selection draws the HI users' and others attention to the impairment.

Definitions As *hearing wish* we define the user's current hearing situation like conversation, listening to music or unintentional hearing. As *cue* we define an element of the user's context such as gestures, activities, mode of locomotion, or location that systematically correlates with a specific hearing wish. *Switching sensitivity* is a HI parameter that determines how fast the HI reacts to a change of the acoustical environment by automatically switching between hearing programs, trading-off stability and plasticity.

Motivation In many situations the acoustical context is similar while the actual hearing wishes diverge. Those situations are currently difficult to distinguish with state-of-the-art HIs only relying on audio analysis. In the following we give an example scenario:

Alice is a typical HI user. Whenever she crosses a busy street she needs her HI to provide her with omnidirectional sound, in particular the sound from approaching cars. Nevertheless, in a similar traffic environment Alice might want to talk to a friend while standing or walking next to a busy road. She now needs her HI to support her with the conversation by suppressing the car noise, e.g. by means of directivity (beam forming to spatially select sound sources).

As the different HI programs represent trade-offs it is crucial to always select an adequate hearing program. During regular meetings with their acousticians HI users usually describe specific situations of their daily life to characterize where the HI failed to adjust properly. In state-of-the-art HIs only sound is available as a user context to let the acoustician tune the situation-specific behavior of the HI. In the example above the acoustician cannot change the HI behavior specifically enough from sound only. A more fine-grained context recognition using additional multimodal sensors could support a specific optimization of the HI in this situation. In particular, additional cues like the user's activity may support the estimation of the user's current hearing wish for the following two cases: (1) if the hearing wish and activity change, but the acoustic context stays the same, and (2) if the hearing wish and user activity stay the same, but the acoustic context changes. Moreover, enhanced context awareness allows using rules such as: "No directivity if using a road!". To not burden the user by calling attention to the hearing impairment an adequate switching sensitivity is important. Nowadays, the switching sensitivity is initially set

[1]State-of-the-art HIs use audio scene analysis to select among up to 4 hearing programs that are optimized for different hearing wishes.

BODYNETS 2011, November 07-08, Beijing, People's Republic of China

DOI 10.4108/icst.bodynets.2011.246911

by the acoustician and does not adapt to the user's context.

HI-BAN Approach We envision to estimate the user's hearing wish from multimodal sensor data to complement state-of-the-art audio analysis. Current HIs already provide wireless connectivity [2]. Hearing instrument body area networks (HI-BANs) represent an emerging trend as they offer a rich source of information complementary to sound [2]. A HI-BAN embeds several sensors and actuators in an on-body-network to sense and provide information through different modalities. Figure 1 illustrates a possible HI-BAN that comprises sound, body and eye movements, location, ambient intelligence, and smart phones with access to the user's agenda and the internet.

Paper Scope and Contributions In this paper, we identify a mapping between the user's context and their current hearing wish. This is currently not known for all situations, because state-of-the-art systems analyze the audio scene only. We analyze which hearing wishes arise in in real world contexts to be recognized with BANs. In addition to estimating the hearing wish we consider switching sensitivity as a temporal characteristic of HI behavior. We show how the identified mapping can be used to improve HIs and derive the sensory requirements for next generations HI-BANs.

2. RELATED WORK

In [11] the authors propose additional on-body sensors to estimate hearing wishes in acoustically ambiguous settings. In a feasibility study they focus on an indoor setting and body and eye movements as additional modalities to sound. The authors found that HI performance benefits from the additional sensor information, especially using acceleration data from the user's head. In [5] an attentive hearing aid based on an eye-tracking device and infrared tags was proposed. Wearers "switch on" selected sound sources such as a person, television or radio by looking at them. Each sound source must be extended an infrared tag so that the HI can focus on the selected sound source. In [8] the authors use an accelerometer integrated into a HI to perform gait analysis. A broad range of methods to use on-body sensors for activity recognition is available from the wearable computing domain [1, 3, 7, 9, 10]. They can be seamlessly translated to recognize similar cues from HI-BANs.

3. REAL-LIFE EXPERIMENT

3.1 Procedure

We designed an experiment to identify the mapping between the user's context and current hearing wish in daily-life situations. The experiment is as close to real life as possible. At the same time we ensure a large number of activities of daily living (ADL) for the sake of statistical relevance of the analysis. A wide range of ADL of elderly people is covered as they form the largest HI user group [6]. Over 4 hours of data have been recorded in real-life settings from a single participant with no hearing impairment. The participant visited places like restaurants, pedestrian areas and crowded tourist sites to cover behavior and conversation in loud noise. The participant passed through safety-critical situations and drove in busy traffic roads by car and bike and walked in a variety of urban situations. A large number of common home situations –typical for elderly people– where included, such as being at home, cooking and eating in the kitchen, listening to the radio, watching TV and a stroll in the park. Since the experiment includes busy places in town, restaurants and public transportation, privacy issues arising from the use of video cameras had to be considered when designing the experiment.

Figure 1: Multimodal HIs: Considering additional modalities to sound, especially for automatic hearing program selection.

3.2 Data Recording and Labeling

The experiment was recorded on video by a second person following the experiment participant with a miniaturized, wearable high definition wide angle camcorder. In a second step, the video was manually annotated offline by the user with their current hearing wish and cues.[2] We used a software designed for annotating videos in multiple parallel tracks. The approach is applicable to arbitrarily defined sets of class labels. We used the following class labels for cues: mode of locomotion (bicycle, car, sit, stand, walk), location (car, inside building, park, pavement, pedestrian area, public transportation, restaurant, street), and gestures (cheers, drinking/eating, indicate direction on bike, open/close door, press button, sudden stop, turnaround, turn head). The labeling approach for the user's current hearing wish comprises *conversation* (for high speech intelligibility also in noisy environments), *orientation* (e.g. when crossing a street), *comfort* (e.g. noise suppression near a construction yard), *music* (for high dynamics), and *unintentional hearing* (no auditory selective attention on any specific sound source). In case of ambiguous hearing wishes the dominating one was selected as a class label. In total the data set comprises more than 1000 class instances.

In addition to the video data for this work, we collected sensor data of body movements with miniaturized inertial measurement units [4], sound from a high-end commercial HI, GPS data, and own-voice data with a throat microphone. We will use the sensor data in future work to implement classifiers for the cues identified as relevant in this work. Figure 2 shows the data recording setup and example situations of the experiment.

4. APPROACH TO IDENTIFY CUES

We analyze the class labels acquired from offline video data labeling to identify cues relevant for hearing wish esti-

[2]We use manual annotation of the cues since this work is about identifying the mapping between cues and hearing wishes. In a deployed system these cues would be detected by means of context recognition techniques. The cues considered here are known from related work to be recognizable in this way.

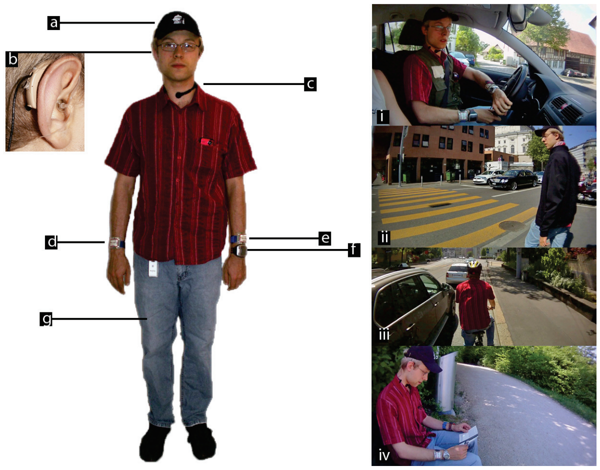

Figure 2: Left: Mobile data recording with miniaturized inertial measurement units at the head (a, hidden under cap), wrists (d, e), and leg (g, hidden under jeans), HI (b), throat microphone (c), and GPS (f). Right: Example situations of the experiment: car drive (i), pedestrian (ii), bicycle ride (iii), and reading (iv).

mation. To identify correlations between hearing wishes and potential cues like gestures, mode of locomotion, or location we calculate histograms to obtain the amount of time (duration) of a hearing wish conditional to the cues. The analysis based on histograms that accumulate the occurrences of the classes is adequate to describe continuous activities such as the mode of locomotion or location of the user. Gestures are treated separately as events as they typically show short durations of less than a second. For gestures we investigate the distribution of hearing wishes that occur after a performed gesture. To also gain insight into the temporal characteristics we calculate the change rate of the hearing wish over time. In this way we can investigate the benefit of adapting the switching sensitivity according to cues.

5. RESULTS AND DISCUSSION

5.1 Mapping of Cues to Hearing Wishes

Gestures Figure 3 shows the distribution of subsequent hearing wishes for different gestures performed by the user. E.g., we found a high probability that the user then participates in a conversation when the user cheers with a glass (100%) or is eating (80%). Moreover, head turns (defined as clearly noticeable left and right turns), sudden stopping, and indication of the direction on the bike correlate strongly with hearing wish orientation (76%, 50%, and 73%, respectively). HIs that take into account user gestures can benefit from a refined automatic hearing program selection. Even anticipation of hearing wishes from gestures is feasible this way, e.g. for the case of picking up the phone or handshaking before a conversation.

Mode of Locomotion and Location Figure 4 shows for each combination of mode of locomotion and location the relative amount of time (duration) of the hearing wish. The normalized hearing wish distributions are visualized in sub blocks of size 3x2 according to the legend in the lower left.

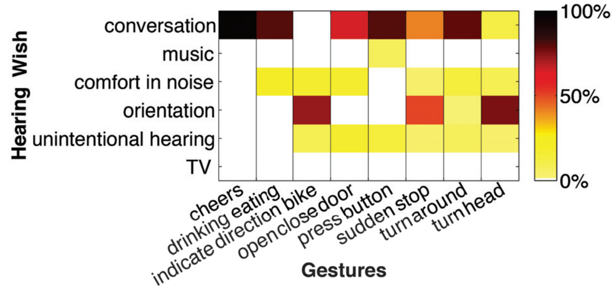

Figure 3: Distribution of subsequent hearing wishes for different gestures performed by the user.

Also the marginal distributions are shown to characterize the correlations for mode of locomotion and location separately. The participant watches TV only when sitting inside a building. When driving a car either conversation or unintentional hearing occurs. For some cases, e.g. for sitting, the additional sensor data does not provide benefit as there is nearly an equal distribution of hearing wishes. However, the combination of modalities allows making a more specific distinction. E.g., sitting in a restaurant is a strong indication for conversation. This represents a plausible behavior of people in restaurants. Restaurant and sitting are not discriminative cues on their own. The identified correlations can be used as a priori-distributions for a refined automatic hearing program selection.

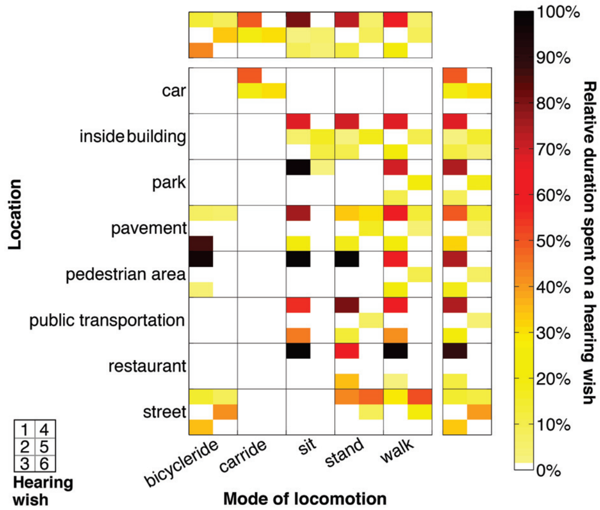

Figure 4: For each combination of mode of locomotion and location the relative amount of time (duration) of a hearing wish. The normalized hearing wish distributions are visualized in sub blocks of size 3x2 according to the legend on the lower left. Hearing wishes comprise (1) conversation, (2) music, (3) comfort in noise, (4) orientation, (5) unintentional hearing, and (6) TV. Also the marginal distributions are shown.

5.2 Adaptive Switching Sensitivity

Figure 5 shows for each combination of mode of locomotion and location the number of changes of hearing wishes per minute calculated from the label time series. For sitting the change rate of 0.67 cpm is relatively low opposed to walking with a relatively high change rate of 1.29 cpm. This is plausible, as activities usually do not change rapidly when sitting. For car ride, bicycle ride and walking on pavement we can observe a relative high hearing wish change rate compared to sit. The identified significant correlations between cues and hearing wish change rate can be used as a dynamic adaption parameter for the switching sensitivity. The HI could adapt to a low or high switching sensitivity. One possible straight-forward implementation of adaptive switching sensitivity is to form two groups of cues for slow and fast switching sensitivity, respectively. The state-of-the-art approach of fixed switching sensitivities is too static to cover all situations over the day as the profile changes during daily life. Thus, introducing two switching sensitivity classes would already be a significant improvement over the state of the art.

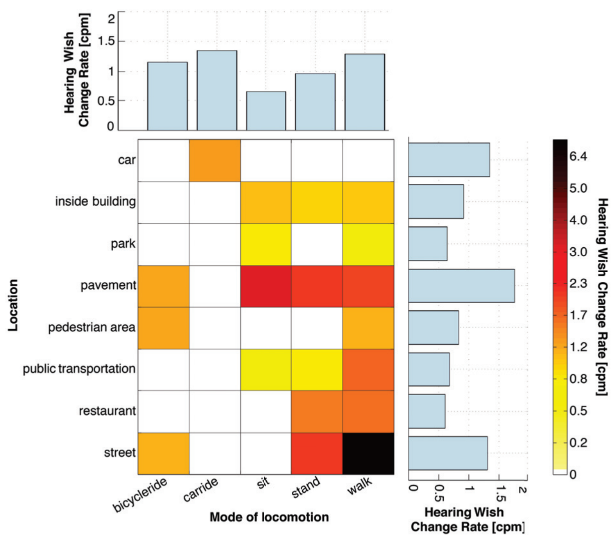

Figure 5: For each combination of mode of locomotion and location the number of changes of hearing wishes per minute is shown. The outer bars indicate the marginal distributions.

5.3 Sensory Requirements to Realize Next Generation of Networked HIs

The mode of locomotion and location were found to be especially important cues. To realize recognition of these cues using the HI-BAN approach a next generation HI must feature a specific set of sensors worn by the user. Based on literature and the relevant cues identified in this paper we can derive the sensory requirements for a next generation HI-BAN to identify these cues: We recommend including at least one accelerometer into the HI [11]. This allows distinguishing sitting and standing from walking [1] and to analyze head movements. Also we recommend a smart phone for location access and GPS data, e.g. to recognize cycling. Optionally a sensor worn in the shoe, like Nike+, could be used to further discriminate cues.

6. CONCLUSION

In this work, we showed how to systematically identify correlations between hearing wishes and user context to enhance automatic hearing program selection if the corresponding cues can be detected with sensors from within the HI-BAN. They can be used to define the a-priori probability distributions of the hearing program selection and to tune the switching sensitivity. Different hearing wishes could be weighted differently, e.g. orientation in traffic might occur rarely, but could be prioritized because of safety issues.
In this one-subject data set, we captured daily activities that are most challenging for state-of-the-art HIs. A larger dataset would allow capturing more events to investigate less frequent situations. We rather aim to cover the most common daily situations which can be detected with a minimal number of additional sensors. This way we minimize power consumption and the burden for the user to wear additional sensors. We nevertheless plan to apply the same method on a larger data set with more users.

Acknowledgement

This work was part funded by CTI project 10698.1 PFLS-LS ”Context Recognition for Hearing Instruments Using Additional Sensor Modalities”.

7. REFERENCES

[1] L. Bao and S. Intille. Activity recognition from user-annotated acceleration data. In *Pervasive Computing*, 2004.

[2] A. Biggins. Benefits of wireless technology. *Hearing Review*, 11 2009.

[3] F. Foerster, M. Smeja, and J. Fahrenberg. Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring. *Computers in Human Behavior*, 15(5):571–583, 1999.

[4] H. Harms et al. ETHOS: Miniature Orientation Sensor for Wearable Human Motion Analysis. In *IEEE Sensors*, 2010.

[5] J. Hart, D. Onceanu, C. Sohn, D. Wightman, and R. Vertegaal. The attentive hearing aid: Eye selection of auditory sources for hearing impaired users. *Human-Computer Interaction –INTERACT*, 2009.

[6] S. Kochkin. MarkeTrak VIII: 25-year trends in the hearing health market. *Hearing Review*, 16(10), 2009.

[7] K. Laerhoven, H. Gellersen, and Y. Malliaris. Long term activity monitoring with a wearable sensor node. In *Wearable and Implantable BSN*, 2006.

[8] B. Lo, J. Pansiot, and G. Yang. Bayesian analysis of sub-plantar ground reaction force with bsn. *2009 Body Sensor Networks*, pages 133–137, 2009.

[9] P. Lukowicz, O. Amft, D. Roggen, and J. Cheng. On-body sensing: From gesture-based input to activity-driven interaction. *Computer*, 43(10), 2010.

[10] S. J. Preece et al. Activity identification using body-mounted sensors—a review of classification techniques. *Physiological Measurement*, 2009.

[11] B. Tessendorf et al. Recognition of hearing needs from body and eye movements to improve hearing instruments. In *International Conference on Pervasive Computing*, 2011.