Recognition of Hearing Needs from Body and Eye Movements to Improve Hearing Instruments

Bernd Tessendorf[1], Andreas Bulling[2], Daniel Roggen[1], Thomas Stiefmeier[1], Manuela Feilner[3], Peter Derleth[3], and Gerhard Tröster[1]

[1] Wearable Computing Lab., ETH Zurich
Gloriastr. 35, 8092 Zurich, Switzerland
{lastname}@ife.ee.ethz.ch
[2] Computer Laboratory, University of Cambridge
15 JJ Thomson Avenue, Cambridge CB3 0FD, United Kingdom
{firstname.lastname}@acm.org
[3] Phonak AG, Laubisrütistrasse 28, 8712 Stäfa, Switzerland
{firstname.lastname}@phonak.com

Abstract. Hearing instruments (HIs) have emerged as true pervasive computers as they continuously adapt the hearing program to the user's context. However, current HIs are not able to distinguish different hearing needs in the same acoustic environment. In this work, we explore how information derived from body and eye movements can be used to improve the recognition of such hearing needs. We conduct an experiment to provoke an acoustic environment in which different hearing needs arise: active conversation and working while colleagues are having a conversation in a noisy office environment. We record body movements on nine body locations, eye movements using electrooculography (EOG), and sound using commercial HIs for eleven participants. Using a support vector machine (SVM) classifier and person-independent training we improve the accuracy of 77% based on sound to an accuracy of 92% using body movements. With a view to a future implementation into a HI we then perform a detailed analysis of the sensors attached to the head. We achieve the best accuracy of 86% using eye movements compared to 84% for head movements. Our work demonstrates the potential of additional sensor modalities for future HIs and motivates to investigate the wider applicability of this approach on further hearing situations and needs.

Keywords: Hearing Instrument, Assistive Technology, Activity Recognition, Electrooculography (EOG).

1 Introduction

Hearing impairment increasingly affects populations worldwide. Today, about 10% of the population in developed countries suffer from hearing problems; in the U.S. even 20% adolescents suffers from hearing loss [22]. Over the last generation, the hearing impaired population grew at a rate of 160% of U.S. population growth [13]. About 25% of these hearing impaired use a hearing instrument (HI) to support them in managing their daily lives [13].

K. Lyons, J. Hightower, and E.M. Huang (Eds.): Pervasive 2011, LNCS 6696, pp. 314–331, 2011.

Over the last decade considerable advances have been achieved in HI technology. HIs are highly specialised pervasive systems that feature extensive processing capabilities, low power consumption, low internal noise, programmability, directional microphones, and digital signal processors [10]. The latest of these systems –such as the Exelia Art by Phonak– automatically select from among four hearing programs. These programs allow the HI to automatically adjust the sound processing to the users' acoustic environment and their current hearing needs. Examples of hearing need support include noise suppression and directionality for conversations in noisy environments.

Satisfying the users' hearing needs in as many different situations as possible is critical. Already a small number of unsupported listening situations causes a significant drop in overall user satisfaction [14]. Despite technological advances current HIs are limited with respect to the type and number of hearing needs they can detect. Accordingly, only 55% of the hearing impaired report of being satisfied with the overall HI performance in common day-to-day listening situations [14]. This is caused, in part, by the fact that adaption is exclusively based on sound. Sound alone does not allow to distinguish different hearing needs if the corresponding acoustic environments are similar. We call this limitation the *acoustic ambiguity problem*.

1.1 Paper Scope and Contributions

In this work we investigate the feasibility of using additional modalities, more specifically body and eye movements, to infer the hearing needs of a person. As a first step toward resolving the acoustic ambiguity problem we focus on one particular listening situation: the distinction between concentrated work while nearby persons have a conversation from active involvement of the user in a conversation. The specific contributions are: 1) the introduction of context-aware HIs that use a multi-modal sensing approach to distinguish between acoustically ambiguous hearing needs; 2) a methodology to infer the hearing need of a person using information derived from body and eye movements; 3) an experiment to systematically investigate the problem of acoustically ambiguous hearing needs in an office environment, and 4) the evaluation of this methodology for automatic hearing program selection.

1.2 Paper Organisation

We first provide an overview of the state-of-the-art in HI technology, introduce the mechanisms that allow HIs to adapt to the user's hearing needs, and discuss the limitations of current systems. We then survey related work and detail our methodology to infer the user's hearing need from body and eye movements. We describe the experiment, discuss its results, and provide a brief outlook on the technical feasibility of integrating body and eye movements into HIs.

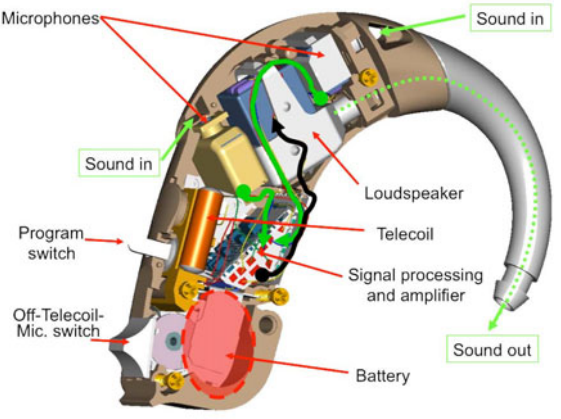

Fig. 1. Components of a modern behind-the-ear (BTE) HI [19]

2 Hearing Instrument Technology

Figure 1 shows the components of a modern behind-the-ear (BTE) HI. HIs are also available in smaller form factors. E.g., Completely-in-the-Canal (CIC) devices can be placed completely inside the user's ear canal. Current systems include a DSP, multiple microphones to enable directivity, a loudspeaker, a telecoil to access an audio induction loop, and a high-capacity battery taking up about a quarter of the HI housing. HIs may also integrate a variety of other accessories such as remote controls, Bluetooth, or FM devices as well as the user's smart phone to form wireless networks, so-called hearing instrument body area networks (HIBANs) [3]. These networking functionalities are part of a rising trend in higher-end HIs. This motivates and supports our investigation of additional sensor modalities for HIs that may eventually be included within the HI itself, or within the wireless network controlled by the HI.

A high-end HI comprises two main processing blocks as shown in Figure 2. The audio processing stages represent the commonly known part of a HI. It performs the traditional audio processing function of the HI and encompasses audio pickup, processing, amplification and playback. The second processing block is the classifier system. It estimates the user's hearing need based on the acoustic environment of the given situation, and adjusts the parameters of the audio processing stages accordingly [12]. The classification is based on spectral and temporal features extracted from the audio signal [4]. The classifier system selects the parameters of the audio processing stages from among a discrete set of parameters known as *hearing programs*. The hearing programs are optimised for different listening situations. Most current high-end HIs distinct four hearing programs: natural, comprehensive hearing (*Speech*), speech intelligibility in noisy environments (*Speech in Noise*), comfort in noisy environments (*Noise*), and listening pleasure for a source with high dynamics (*Music*). The hearing programs represent trade-offs, e.g. speech intelligibility versus naturalness of sound,

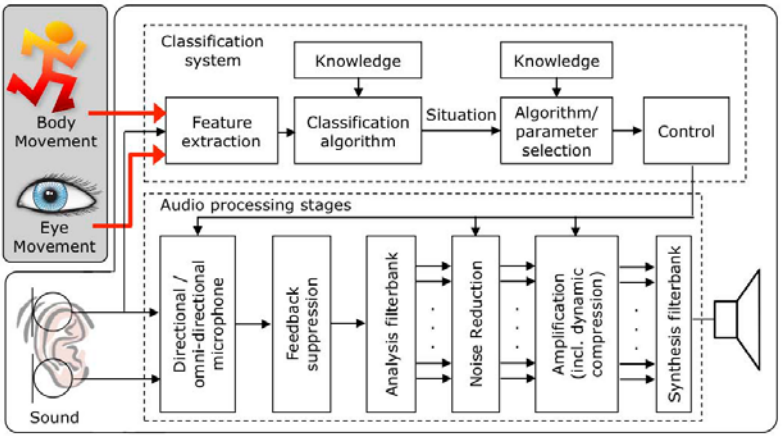

Fig. 2. Bottom: audio processing stages of the HI, from microphone pick-up to amplified and processed sound playback. Top right: classification of the acoustic environment based on sound to adjust the parameters of the audio processing stages. Top left: the extension proposed in this paper. Body and eye movement data are included in the classification system to select the appropriate hearing program. (Figure extended from [10]).

or omnidirectional listening versus directivity. The automatic program selection allows the hearing impaired to use the device with no or only a few manual interactions such as program change and volume adjustment. Adaptive HIs avoid drawing attention to the user's hearing deficits. Users consider automatic adaption mechanisms as useful [4]. Further technical details on HI technology can be found in [10, 21].

2.1 The Acoustic Ambiguity Problem

HIs select the most suitable hearing program according to the user's acoustic environment. The current acoustic environment is used as a proxy for the user's actual hearing need. This approach works well as long as the acoustic environment and hearing need are directly related. This assumption does not hold in all cases and leads to a limitation we call the *acoustic ambiguity problem*: Specifically, in the same acoustic environment a user can have different hearing needs that require different hearing programs. A sound-based adaption mechanism cannot distinguish between these different hearing needs. Therefore, it is important to not only analyze the acoustic environment but to also assess the relevance of auditory objects [23]. The challenge here is not the effectiveness of the dedicated hearing programs but rather automatically adapting the hearing program to the specific *hearing need*, rather than to the *acoustic environment*. The following story illustrates the acoustic ambiguity problem:

Alice suffers from hearing impairment and works in an open office space. Alice is focused on her assigned task when Bob enters the office space to talk to a

colleague sitting next to Alice. Alice's HI senses speech in noise and optimizes for speech intelligibility. She now has a hard time focussing on her work, as the HI adapts to the distracting conversation that occurs around her. Then Bob starts talking to Alice. She now needs the HI to support her interaction with colleagues in the noisy environment. Alice doesn't like to select hearing programs manually and desires a robust automatic adaption to her current hearing need.

In the first case, the HI user takes part in a conversation, in the second case, the user could be concentrated on her work and and experiences the conversation as noise. The key challenge in this example is to assess the relevance of speech in the acoustic environment to the HI user. The HI needs to choose between a hearing program that optimizes speech intelligibility and a hearing program treating the speech as noise for user comfort. In both situations, the HI detects the same acoustic environment and thus cannot select a suitable hearing program in both of the cases. A possible strategy is a static "best guess" choice based on a predefined heuristic rule. It could favor speech intelligibility over comfort in noise as social interaction is generally considered important.

Other typical situations in which state of the art classification systems fail include listening to music from the car radio while driving or conversing in a cafe with background music [10].

2.2 Vision of a Future HI

We envision the use of additional modalities to distinguish between ambiguous hearing need requirements in the same acoustic environment. These modalities will be included within the HI itself, or within a wireless network controlled by the HI. Wireless networking functionalities are now starting to appear in higher-end HIs. These new sensors need not be specifically deployed for HIs: they may be shared with other assistive technologies, such as systems designed to detect falls or to monitor physiological parameters. Thus, we see the HI as one element included in a broader set of ambient assisted living technologies. Wearable and textile integrated sensors have become available and sensor data from a mobile phone that may be carried by an individual can be used. We believe the next step in HI technology is to utilize this infrastructure to improve HI performance.

3 Related Work

Various sensor modalities have been proposed to detect social interaction, conversation, or focus of attention from wearable sensors. In [8] body-worn IR transmitters were used to measure face-to-face interactions between people with the goal to model human networks. All partners involved in the interaction needed to wear a dedicated device.

In [11] an attentive hearing aid based on an eye-tracking device and infrared tags was proposed. Wearers should be enabled to "switch on" selected sound sources such as a person, television or radio by looking at them. The sound source needed to be attached with a device that catched the attention of the

hearing aid's wearer so that only the communication coming from the sound source was heard.

In [7] different office activities were recognised from eye movements recorded using Electrooculography with an average precision of 76.1% and recall of 70.5%: copying a text between two screens, reading a printed paper, taking hand-written notes, watching a video, and browsing the web. For recognising reading activity in different mobile daily life settings the methodology was extended to combine information derived from head and eye movements [6].

In [18] a vision-based head gesture recognizer was presented. Their work was motivated by the finding that head pose and gesture offer key conversational grounding cues and are used extensively in face-to-face interaction among people. Their goal was to equip an embodied conversational agent with the ability to perform visual feedback recognition in the same way humans do. In [9] the kinematic properties of listeners' head movements were investigated. They found a relation of timing, tempo and synchrony movements of responses to conversational functions.

Several researchers investigated the problem of detecting head movements using body-worn and ambient sensors. In [1] an accelerometer was placed inside HI-shaped housing and worn behind the ear to perform gait analysis. However, the system was not utilised to improve HI behavior.

Capturing the user's auditory selective attention helps to recognise a person's current hearing need. Research in the field of electrophysiology focuses on mechanisms of auditory selective attention inside the brain [24]. Under investigation are event-related brain potentials using electroencephalography (EEG). In [17] the influence of auditory selection on the heart rate was investigated. However, the proposed methods are not robust enough yet to distinguish between hearing needs and are not ready yet for deployment in mobile settings.

All these approaches did not consider sensor modalities which may be included in HIs, or assumed the instrumentation of all participants in the social interactions. In [26], analysis of eye movements was found to be promising to distinguish between working and interaction. Head movements were found to be promising to detect whether a person is walking alone or walking while having a conversation. However, the benefit of combining modalities was not investigated. Moreover, the actual improvement in hearing program selection based on the context recognition was not shown.

4 Experiment

4.1 Procedure

The experiment in this work was designed to systematically investigate acoustically ambiguous hearing needs in a reproducible and controllable way, still remaining as naturalistic as possible. We collected data from 11 participants (six male, five female) aged between 24 and 59 years, recruited from within the lab. The participants were normal hearing and right handed without any known attributes that could impact the results.

Table 1. Experimental procedure to cover different listening situations and hearing needs. The procedure was repeated eight times with the different office activities mentioned above and with the participants being seated and standing.

Time Slot [min]	Situation	Hearing Need
1	Participant and colleague are working	work
2	Disturber and colleague converse	work
3	Disturber and participant converse	conversation
4	Disturber and colleague converse	work
5	Colleague and participant converse	conversation

The experiment took place in a real but quiet office room. The participant and an office colleague worked in this office. A third person, the disturber, entered the office from time to time to involve them in a conversation. The participants were given tasks of three typical office activities: Reading a book, writing on a sheet of paper, typing text on a computer. The participants had no active part in controlling the course of events. They were instructed to focus on carrying out their given tasks and to react naturally. This assures that the resulting body and eye movements are representative for these activities.

The experiment was split in one minute time slots each representing a different situation and hearing need (see Table 1). In the first minute, the participant worked concentrated on his task. In the second minute, the participant tried to stay concentrated while the office colleague was talking to the disturber. In the third minute, the participant was interrupted and engaged in a conversation with the disturber. In the fourth minute, the disturber talked to the colleague again. In the fifth minute, the participant and the colleague had a conversation.

This procedure was repeated eight times with the office activities mentioned above and with the participants being seated and standing. The total experiment time for each participant was about 1 hour. We then assigned each of these activities to one of the following two hearing needs.

Conversation includes situations in which the participant is having a conversation. The HI is supposed to optimize for speech intelligibility, i.e. the hearing program should be "Speech in Noise" throughout. Figure 3 shows all four combinations of sitting and standing while talking to the conversation partners.

Work includes situations in which the participant is carrying out a work task. This case covers situations in which no conversation is taking place around him and situations in which two colleagues are having a conversation the participant is not interested in. The HI is supposed to be in a noise suppression program called "Noise". Figure 4 shows the participant work sitting and standing. Figure 5 shows the participant work in speech noise for the sitting case only.

4.2 Performance Evaluation

We investigate how accurate we can distinguish the two classes *conversation* and *work*. The hearing programs we declared as optimal for each of the situations

Fig. 3. Situations with the hearing need *Conversation*, including all four combinations of sitting and standing conversation partners. The HI is supposed to be in a program optimizing for speech intelligibility (*Speech In Noise*).

Fig. 4. Situations with the hearing need *Noise* for the case *Work*. Working tasks include reading a book, writing on a sheet of paper, and typing a text on the computer. The participant works sitting and standing.

Fig. 5. Situations with the hearing need *Noise* for the case *Work in Speech Noise*. The participant tries to focus on his working task while two colleagues are having a conversation. Only the sitting case is shown here.

served as ground truth: *Speech In Noise* for conversation and *Noise* for work. It is important to note that the *Noise* program is not optimized for supporting the user with concentrated work, but is the best choice among the available hearing programs in our conversation cases. Robust detection of working situations would enable to augment existing HIs with a dedicated program and sound signal processing strategies. For evaluation we compare for each signal window whether the classification result corresponds to the ground truth. We count how often classification and ground truth match in this two-class problem to obtain an accuracy value. In addition, we obtained as a baseline the classification result based on sound. To this end, we analysed the debug output of an engineering sample of a HI[1].

4.3 Data Collection

For recording body movements we used an extended version of the Motion Jacket [25]. The system features nine MTx sensor nodes from *Xsens Technologies* each comprising a 3-axis accelerometer, a 3-axis magnetic field sensor, and a 3-axis gyroscope. The sensors were attached to the head, the left and right upper and lower arms, the back of both hands, and the left leg. The sensors were connected to two XBus Masters placed in a pocket at the participants' lower back. The sampling rate is 32 Hz.

For recording eye movements we chose Electrooculography (EOG) as an inexpensive method for mobile eye movement recordings; it is computationally light-weight and can be implemented using on-body sensors [5]. We used the Mobi system from Twente Medical Systems International (TMSI). The device records a four-channel EOG with a joint sampling rate of 128 Hz. The participant wore it on a belt around the waist as shown in Figure 6. The EOG data was collected using an array of five electrodes positioned around the right eye as shown in Figure 6. The electrodes used were the 24mm Ag/AgCl wet ARBO type from Tyco Healthcare equipped with an adhesive brim to stick them to the skin. The horizontal signal was collected using two electrodes on the edges of both eye sockets. The vertical signal was collected using one electrode above the eyebrow and another on the lower edge of the eye socket. The fifth electrode, the signal reference, was placed away from the other electrodes in the middle of the forehead. Eye movement data was saved together with body movement data on a netbook in the backpack worn by the participant.

We used two Exelia Art 2009 HIs from Phonak worn behind the left and the right ear. For the experiment we modified the HIs to use them for recording only the raw audio data rather than logging the classification output in real-time. With the raw audio data the HI behavior in the conducted experiment can be reconstructed offline. Using the same noise for overlay gives equal conditions for each participant to rule out different background noise as an effect on the resulting performance. Moreover, it is possible to simulate for different acoustic environments, e.g. by overlaying office noise. Another advantage of recording

[1] This work was carried out in collaboration with a hearing instrument company.

Fig. 6. Sensor setup consisting of HIs (1), a throat microphone (2), an audio recorder (3), five EOG electrodes (h: horizontal, v: vertical, r: reference), as well as the Xsens motion sensors placed on the head (4a), the upper (4b) and lower (4c) arms, the back of both hands (4d), the left leg (4e), two XBus Masters (4d), and the backpack for the netbook (5)

raw audio data is the possibility to simulate the behavior with future generation of HIs. We used a portable audio recorder from SoundDevices to capture audio data with 24 bit at 48 kHz. Although not used in this work, participants also wore a throat microphone recording a fifth audio channel with 8 bit at 8 kHz. Based on both sound recordings we investigate detection of active conversation based on own-speech detection in future research.

Data recording and synchronisation was handled using the Context Recognition Network (CRN) Toolbox [2]. We also videotaped the whole experiment to label and verify the synchronicity of the data streams.

5 Methods

5.1 Analysis of Body Movements

We extract features on a sliding window on the raw data streams from the 3-axis accelerometers, gyroscopes and magnetometers. For the magnetometer data we calculate mean, variance, mean crossing and zero crossing. For the gyroscope data we additionally extract the rate of peaks in the signal. For the accelerometers data we calculate the magnitude based on all three axes. Based on a

parameter sweep we selected a window size of 3 seconds and a step size of 0.5 second for feature extraction.

5.2 Analysis of Eye Movements

EOG signals were first processed to remove baseline drift and noise that might hamper eye movement analysis. Afterwards, three different eye movement types were detected from the processed EOG signals: saccades, fixations, and blinks. All parameters of the saccade, fixation, and blink detection algorithms were fixed to values common to all participants. The eye movements returned by the detection algorithms were the basis for extracting different eye movement features using a sliding window. Based on a parameter sweep we set the window size to 10 seconds and the step size to 1 second (a more detailed description is outside the scope of this paper but can be found in [7]).

5.3 Feature Selection and Classification

The most relevant features extracted from body and eye movements were selected with the maximum relevance and minimum redundancy (mRMR) method [20]. Classification was done using a linear support vector machine (see [15] for the specific implementation we used). We set the penalty parameter to $C = 1$ and the tolerance of termination criterion to $\epsilon = 0.1$. Classification and feature selection were evaluated using a leave-one-participant-out cross-validation scheme. The resulting train and test sets were standardised to have zero mean and a standard deviation of one. Feature selection was performed solely on the training set.

5.4 Analysis of Sound

We used the classification output of commercial HIs as a baseline performance for sound based classification. We electrically fed the recorded audio stream described in section 4.3 back into HIs and obtained the selected hearing programs over time with a sampling rate of 10 Hz. To simulate a busy office situation we overlaid the recorded raw audio data with typical office background noise. In silent acoustic environments without noise, the hearing instrument remains mainly in the *Clean Speech* program for both the working and the conversation situation. We focus on the scenario with noise: The HI needs to decide wether optimizing for speech is adequate or not.

5.5 Data Fusion

To combine the unimodal information from the different motion sensors we used a fusion approach on feature-level. We built a feature vector comprising features from each of the sensors. To combine the multimodal information from body movement, eye movement, and sound we used majority voting as a standard fusion method on classifier-level. When there was no majority to make a decision we repeated the most recent decision. In this way, we suppress hearing program changes based on low confidence.

6 Results and Discussion

6.1 Analysis of the Different Modalities

We first evaluated the performance of the different modalities. Accuracies are given for the two-class classification problem comprising active conversation and working while colleagues are having a conversation. Figure 7 shows the accuracies for distinguishing the hearing needs using sound, body movements, eye movements, and combinations of these modalities averaged over all participants. The results for body movements are based on sensors attached to all nine body locations whereas the results for sound-based adaption are based on the classification output of the HIs.

The limited recognition accuracy of 77% for adaption based on sound is a consequence of the acoustic ambiguity problem that has been provoked in this scenario. The sound based analysis does not distinguish between relevant and irrelevant speech. The HI optimizes for speech in both of the cases described in section 4.1: When the participant is having a conversation and also when the colleagues are having a conversation.

As can be seen from Figure 7, recognition based on body movement data from all available movement sensors (placed at head, back, arms, hands, leg) achieves the best performance with an accuracy of 92%. Adaptation based on eye movement performs slightly worse with 86% accuracy. Looking at combinations of different modalities shows that the joint analysis of body and eye movements has an accuracy of 91%, sound and body movement results in 90% accuracy, and combination of sound and eye movements yields 85% accuracy. Complementing body movements with eye movements or sound results in a lower standard standard deviation, meaning more robustness across different users. First results suggest the inclusion of movement sensors additionally to sound into the HI.

6.2 Analysis of Body Locations

Based on these findings we selected body movements for further analysis. We investigated on which body locations the movement sensors provided the highest

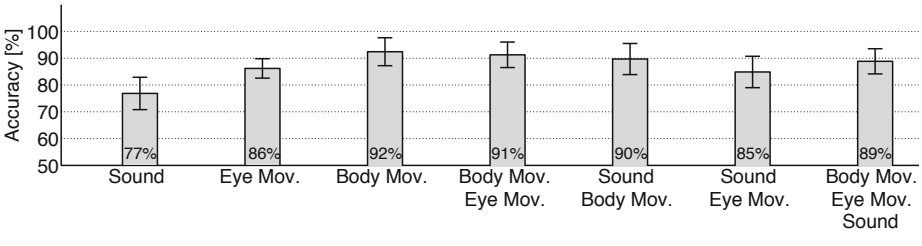

Fig. 7. Accuracies for distinguishing the hearing needs in our scenario based on sound, eye movements, body movements (placed at head, back, arms, hands, leg), and all possible combinations. Results are averaged over all participants with the standard deviation indicated with black lines.

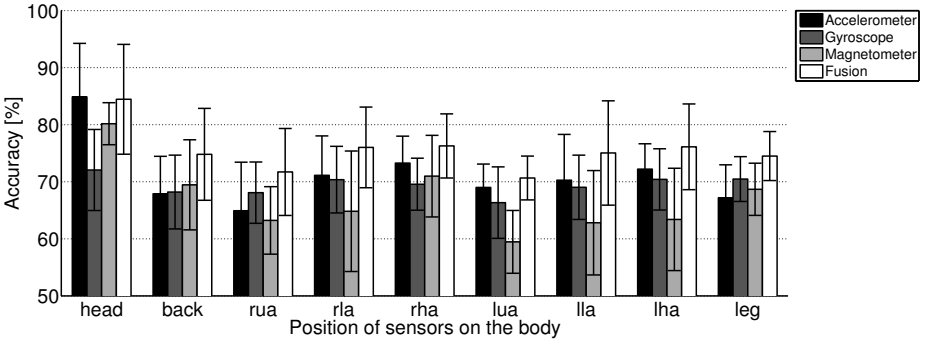

Fig. 8. Accuracies of adaption based on individual sensors (Accelerometer, Magnetometer, Gyroscope) for each of the 9 locations on the body: head, back, left upper arm (lua), left lower arm (lla), left hand (lha), right upper arm (rua), right lower arm (rla), right hand (rha), and leg. Results are averaged over all participants with the standard deviation indicated with black lines.

accuracies. Figure 8 provides the accuracies for adaption using individual sensors (Accelerometer, Magnetometer, Gyroscope) as well as using sensor fusion on feature level for each of the nine body locations averaged over all participants.

Figure 8 shows that from all individual body locations the sensor on the head yields the highest performance with accuracies between 72% for the gyroscope and 85% for the accelerometer. It is interesting to note that fusing the information derived from all three sensors types at the head does not further improve recognition performance (see first group of bars in Figure 8). For all other body location, sensor fusion consistently yields the best recognition performance. Single sensors placed on other body locations perform considerably worse with accuracies ranging from 59% (magnetometer on the left upper arm, lua) to 73% (accelerometer on the right hand, rha). These sensors may still prove beneficial if combined with other sensors located at other parts of the body. The higher utility of analyzing sensors on the right arm (rha, rla, rua) can be explained by the fact that all participants were right handed.

6.3 Further Analysis of the Head Location

As shown in the previous sections, the head is the most relevant individual body location. The sensors at this location are also the most promising with respect to a later implementation into a HI.

Figure 9 shows the accuracies for distinguishing the hearing needs based on sound, head movements, eye movements, and all possible combinations. As can be seen from Figure 9, from the three individual modalities, an accuracy of 86% was achieved using eye movements. Moreover, the standard deviation is lower than the one for head movements that yields an accuracy of 84%. From all individual modalities, eye movement analysis performs best. From all combinations

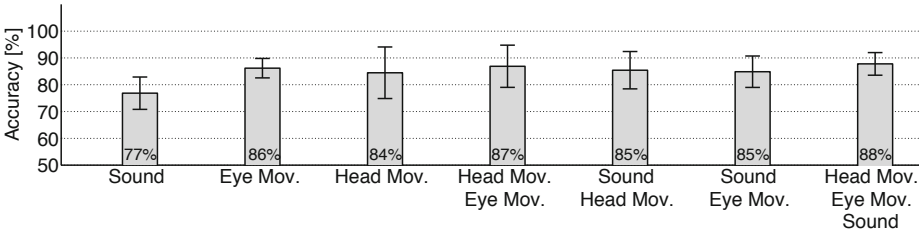

Fig. 9. Accuracies for distinguishing the hearing needs in our scenario based on sound, eye movements, head movements, and all possible combinations. Results are averaged over all participants with the standard deviation indicated with black lines.

of two modalities (bars 4–6 in Figure 9), joint analysis of head and eye movements perform with 87%. The combination of all three modalities yields the highest accuracy of 88%.

Adaption based on eye movements (86%) outperforms adaption based on head movements (84%). As described in section 5, eye movement analysis requires a three times larger data window size (10 seconds) than body movement analysis (3 seconds), leading to a larger classification latency. The joint analysis of body and eye movements combines the more long-term eye movement analysis, and more short-term body movements and yields an accuracy of (85%).

Taking into account movement from all body location corresponds to the idea of leveraging the HIBAN described in section 2. Sensing head and eye movements corresponds to the idea to eventually integrate all sensors into the HI. The HIBAN approach leads to higher than the stand-alone approach at the cost of additional locations on the body that have to be attached with a sensor. The two cases represent a trade-off between accuracy and required number of body locations attached with sensors. Hearing impaired can decide to take the burden of wearing additional sensors to benefit from better hearing comfort. Besides, smartphone and on-body sensors are more and more likely to be available. As shown, the system functions stand-alone with reduced performance.

6.4 Individual Results for Each Participant

To further investigate the large standard deviation for head movements we additionally analysed the individual recognition performance for each participant. Figure 10 shows the accuracy of choosing the correct program for adaption based on sound, head movements, eye movements, and their fusion on feature level for each individual participant. This analsyis reveals that for four participants eye movements performed best, for the remaining 7 participants head movements performed best. Eye movements provide more consistent high performances for all participants between 82% and 91%. Results for head movements were less consistent. In particular participant 9 and 10 showed reduced accuracies of 61% and 74%. A possible reason for this can be a displaced sensor, e.g. caused by the user adjusting the cap. For eye movements the variability is smaller in the given

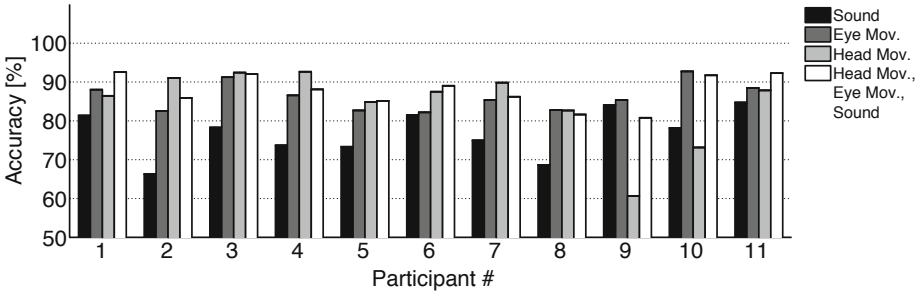

Fig. 10. Accuracy for adaption based on sound, head movements, eye movements and their combination, individually for each participant

data set. Sound adaption compares worse to body and eye movement adaption since this scenario intentionally contains acoustically ambiguous hearing needs.

6.5 Integrating Body and Eye Movement Sensing into a HI

The integration of additional sensor modalities is within reach of future HIs driven by the trend of HIBANs. HIs may in some cases be one of many deployed ambient assisted living technologies. Thus, wearable and textile integrated sensors, as well as the user's smart-phone may become part of the HIBAN. Future HIs can also take advantage of additional sensors that are already deployed for other purposes (e.g. motion sensing for fall detection). This reduces the user burden of utilizing multiple sensors while improving his auditive comfort.

Whenever HiBANs are not available, sensors could also be completely integrated into the HI itself to provide a stand-alone solution. Low power accelerometers with small footprints are available for integration into a HI. EOG is an inexpensive method for mobile eye movement recording. These characteristics are crucial for future integration of long-term eye movement data into future HIs in mobile daily life settings. EOG integration into a HI could follow integration achievements of EOG into glasses [7] or headphones [16].

6.6 Limitations

Although we significantly enhanced the distinction of two ambiguous auditory situations, our multimodal context recognition approach remains a proxy to infer what is essentially a subjective matter: the *subjective hearing need* of a person. Thus, even a perfect context recognition would not guarantee that the hearing need is detected correctly all the time. Ultimately, this would require capturing the user's auditory selective attention. Our evaluation is based on the recognition accuracy compared to the objective ground truth defined in the scenario. However, to assess the actual benefit experienced by the user, a more thorough user study with hearing impaired will need to be carried out.

We currently investigated a single ambiguous auditory situation. Nevertheless there are a large number of other ambiguous situations for current hearing instruments. Our objective is to identify the smallest subset of additional sensor modalities which can help to distinguish a wide range of currently challenging auditory situations. Thus, this work in an office scenario is an exemplary proof-of-concept approach. It still needs to be shown that the approach can be generalised and that one can resolve ambiguity in a sufficient number of other situations to justify the inclusion of additional sensor modalities within HIs.

The office scenario we chose may be a limitation. We chose a specific office work situation, but a variety of other office situations are thinkable, e.g. with more conversation partners and different activities. For this proof-of-concept study it was necessary to choose a trade-off between variety and control to collect data in a reproducible manner for multiple participants. After the experiment we went through a short questionnaire with each participant. The general feedback was, that the sensor equipment was found to be bulky, but overall the participants felt that they were not hindered to act natural.

Overlaying background noise as described in section 5.4 may be a limitation. We overlaid one typical office background noise. Many different kinds and intensities are thinkable. In some cases, the performance of the sound-based HI might be better. However, the performance based on body and eye movement is independent of the present sound. As a further potential limitation the participants may not act the same as they would if there is actual background noise.

6.7 Considerations for Future Work

There are a large number of other challenging situations that are faced by current HIs, e.g. listening to music from the car radio while driving, reading a book in a busy train, or conversing in a cafe with background music. This motivates the investigation of additional modalities, acoustic environments, and hearing situations in future work. A critical issue will be the trade-off in improving context awareness in HIs while minimising the burden caused by additional sensors. Possible additional sensor modalities are the user's current location, proximity information, or information from other HIs or the environment. Based on the promising results achieved in our proof-of-concept study, we plan to deploy our system in further real-life outdoor scenarios to study the benefit in everyday life experienced by the user.

7 Conclusion

Hearing instruments have emerged as true pervasive computers and are fully integrated into their user's daily life. In this work we have shown that multimodal fusion of information derived from body and eye movements is a promising approach to distinguish acoustic environments that are challenging for current hearing instruments. These results are particularly appealing as both modalities can potentially be miniaturised and integrated into future HIs.

Acknowledgments. This work was part funded by CTI project 10698.1 PFLS-LS "Context Recognition for Hearing Instruments Using Additional Sensor Modalities". The authors gratefully thank all participants of the experiment and the reviewers for their valuable comments.

References

1. Atallah, L., Aziz, O., Lo, B., Yang, G.Z.: Detecting walking gait impairment with an ear-worn sensor. In: International Workshop on Wearable and Implantable Body Sensor Networks, pp. 175–180 (2009)
2. Bannach, D., Amft, O., Lukowicz, P.: Rapid prototyping of activity recognition applications. IEEE Pervasive Computing 7, 22–31 (2008)
3. Biggins, A.: Benefits of wireless technology. Hearing Review (2009)
4. Büchler, M., Allegro, S., Launer, S., Dillier, N.: Sound Classification in Hearing Aids Inspired by Auditory Scene Analysis. EURASIP Journal on Applied Signal Processing 18, 2991–3002 (2005)
5. Bulling, A., Roggen, D., Tröster, G.: Wearable EOG goggles: Seamless sensing and context-awareness in everyday environments. Journal of Ambient Intelligence and Smart Environments 1(2), 157–171 (2009)
6. Bulling, A., Ward, J.A., Gellersen, H.: Multi-Modal Recognition of Reading Activity in Transit Using Body-Worn Sensors. ACM Transactions on Applied Perception (to appear, 2011)
7. Bulling, A., Ward, J.A., Gellersen, H., Tröster, G.: Eye Movement Analysis for Activity Recognition Using Electrooculography. IEEE Transactions on Pattern Analysis and Machine Intelligence 33(4), 741–753 (2011)
8. Choudhury, T., Pentland, A.: Sensing and modeling human networks using the sociometer. In: ISWC, p. 216. IEEE Computer Society, Washington, DC, USA (2003)
9. Hadar, U., Steiner, T.J., Clifford Rose, F.: Head movement during listening turns in conversation. Journal of Nonverbal Behavior 9(4), 214–228 (1985)
10. Hamacher, V., Chalupper, J., Eggers, J., Fischer, E., Kornagel, U., Puder, H., Rass, U.: Signal processing in high-end hearing aids: State of the art, challenges, and future trends. EURASIP Journal on Applied Signal Processing 18(2005), 2915–2929 (2005)
11. Hart, J., Onceanu, D., Sohn, C., Wightman, D., Vertegaal, R.: The attentive hearing aid: Eye selection of auditory sources for hearing impaired users. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5726, pp. 19–35. Springer, Heidelberg (2009)
12. Keidser, G.: Many factors are involved in optimizing environmentally adaptive hearing aids. The Hearing Journal 62(1), 26 (2009)
13. Kochkin, S.: MarkeTrak VIII: 25-year trends in the hearing health market. Hearing Review 16(11), 12–31 (2009)
14. Kochkin, S.: MarkeTrak VIII: Consumer satisfaction with hearing aids is slowly increasing. The Hearing Journal 63(1), 19 (2010)
15. Lin, C.J.: LIBLINEAR - a library for large linear classification (February 2008) http://www.csientuedutw/~cjlin/liblinear/
16. Manabe, H., Fukumoto, M.: Full-time wearable headphone-type gaze detector. In: Ext. Abstracts of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1073–1078. ACM Press, New York (2006)

17. Molen, M., Somsen, R., Jennings, J.: Does the heart know what the ears hear? A heart rate analysis of auditory selective attention. Psychophysiology (1996)
18. Morency, L.P., Sidner, C., Lee, C., Darrell, T.: Contextual recognition of head gestures. In: ICMI 2005: Proceedings of the 7th International Conference on Multimodal Interfaces, pp. 18–24. ACM, New York (2005)
19. Naylor, G.: Modern hearing aids and future development trends, http://www.lifesci.sussex.ac.uk/home/Chris_Darwin/BSMS/Hearing%20Aids/Naylor.ppt
20. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(8) (2005)
21. Schaub, A.: Digital Hearing Aids. Thieme Medical Pub. (2008)
22. Shargorodsky, J., Curhan, S., Curhan, G., Eavey, R.: Change in Prevalence of Hearing Loss in US Adolescents. JAMA 304(7), 772 (2010)
23. Shinn-Cunningham, B.: I want to party, but my hearing aids won't let me? Hearing Journal 62, 10–13 (2009)
24. Shinn-Cunningham, B., Best, V.: Selective attention in normal and impaired hearing. Trends in Amplification 12(4), 283 (2008)
25. Stiefmeier, T., Roggen, D., Ogris, G., Lukowicz, P., Tröster, G.: Wearable activity tracking in car manufacturing. IEEE Pervasive Computing 7(2), 42–50 (2008)
26. Tessendorf, B., Bulling, A., Roggen, D., Stiefmeier, T., Tröster, G., Feilner, M., Derleth, P.: Towards multi-modal context recognition for hearing instruments. In: Proc. of the International Symposium on Wearable Computers (ISWC) (2010)