34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Ear-Worn Reference Data Collection and Annotation for Multimodal Context-Aware Hearing Instruments

Bernd Tessendorf, Peter Derleth, Manuela Feilner, Franz Gravenhorst, Andreas Kettner, Daniel Roggen, Thomas Stiefmeier, Gerhard Tröster

Abstract— **In this work we present a newly developed ear-worn sensing and annotation device to unobtrusively capture head movements in real life situations. It has been designed in the context of developing multimodal hearing instruments (HIs), but is not limited to this application domain. The ear-worn device captures triaxial acceleration, rate of turn and magnetic field and features a one-button-approach for real-time data annotation through the user. The system runtime is over 5 hours at a sampling rate of 128 Hz. In a user study with 21 participants the device was perceived as comfortable and showed a robust hold at the ear. On the example of head acceleration data we perform unsupervised clustering to demonstrate the benefit of head movements for multimodal HIs. We believe the novel technology will help to push the boundaries of HI technology.**

I. INTRODUCTION

A major trend in hearing instrument (HI) technology is towards multimodal sensing in addition to sound to improve automatic hearing program selection [1], [2]. In this context the challenge arises to collect and annotate multimodal reference data to identify hearing situations to be improved and to train the multimodal classifier running within the HI. To address this need we developed in tight collaboration with a HI manufacturer, a HI acoustician and HI users an ear-worn multimodal data collection and annotation device that meets the following requirements:

- providing *additional sensors to sound* that are subject to investigation for potential integration in future HIs,
- *long-term deployment in real life settings* to cover situations that cannot be represented in laboratory environments,
- *unobtrusiveness* as the system should influence the user's behavior as little as possible,
- *data annotation solution* as annotated data is required for applying machine learning algorithms and statistical analyses.

We place the device at the user's ear, because this location is especially appealing:

- HIs are worn at the same location, thus data from our sensors corresponds to the data from sensors that will eventually be integrated into future HIs,
- in previous work the authors found head movements as promising to distinguish hearing wishes [1],
- the location behind the ear is practical known from the widespread of use HIs

Our device integrates three MEMS sensors: a triaxial accelerometer, a triaxial gyroscope, and a triaxial magnetic field sensor. The variety of sensors allows us to determine the most discriminant subset of sensors for a given hearing situation. Minimizing the required subset of sensors is crucial, because integration of hardware and algorithms into HIs imposes strong volume and power restrictions.

A. Paper Scope and Contributions

We present a newly developed ear-worn sensing and annotation device to unobtrusively capture head movements in real life situations. We characterize the system from the technical and user's point of view. On the example of head acceleration data we perform unsupervised clustering to demonstrate the benefit of head movement data for multimodal HIs.

B. Paper Organization

We first provide an overview of the state of the art in related data recording and annotation devices and discuss the limitations of current systems. We then describe and characterize our proposed system. Finally, we discuss the findings from the unsupervised clustering of head acceleration data, conclude, and give an outlook.

II. RELATED WORK

A. Ear-Worn Data Recording Devices

In [3] the authors present an ear-worn accelerometer for activity recognition. However, it does not offer data annotation functionalities or additional sensors. In [4] the authors present an ear-worn device for wearable photoplethysmography. A triaxial accelerometer is used to implement motion-adaptive noise cancellation filters for the physiological signal. Another ear-worn photoplethysmography sensor is presented in [6]. The device does not feature accelerometers but a dedicated flexible transducer structure, which can adapt to a variety of skin surface contours. In [5] the authors use an ear-worn camera to detect 3D hand gestures. Feedback is given acoustically by earphones and visually by a miniaturized projector. Unobtrusive operation was not a design goal for both the system interaction modalities and the housing.

To the best of our knowledge no ear-worn systems are currently available that include a triaxial accelerometer, gyroscope and magnetic field sensor. Most existing devices are too bulky to be worn unobtrusively behind the ear and do not offer any data annotation functionality.

B. Tessendorf, F. Gravenhorst, A. Kettner, D. Roggen, T. Stiefmeier, and G. Tröster are with the Wearable Computing Lab., ETH Zurich, Gloriastrasse 35, 8092 Zurich, Switzerland. tessendorf@ife.ee.ethz.ch P. Derleth and M. Feilner are with Phonak AG, Laubisrütistrasse 28, 8712 Stäfa, Switzerland.

978-1-4577-1787-1/12/$26.00 ©2012 IEEE

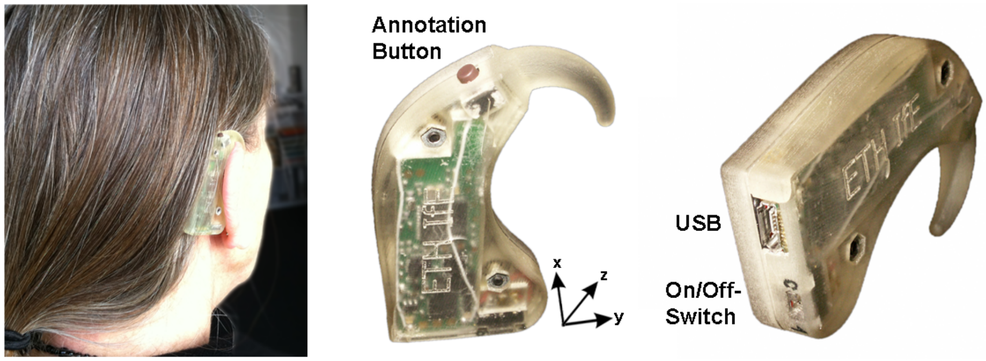

Fig. 1. The newly developed ear-worn sensing (triaxial accelerometer, gyroscope and magnetic field sensor) and annotation device: Shown are a user wearing the device, and the device with the annotation button and the USB connector for charging the battery and downloading the recorded data.

B. Data Annotation

Designing an unobtrusive wearable data annotation system is a well-known challenge in the wearable computing research community. Manual offline annotation is a cumbersome, error prone, and extremely time consuming process [7]. Accurate but time consuming video annotation can be used to annotate a subset of the data set, and then machine learning techniques are used to make a "draft annotation" of the remainder of the dataset. This saves time for the experimenter, as he then only needs to skim through the draft annotations in places where there is a low confidence in the automatic annotation. However, a video camera needs to be used, which is more obtrusive and even problematic in public space due to privacy issues. Semi- or unsupervised algorithms as well as transfer learning [8] can be used to mitigate the annotation problem. However, this way no quantitative evaluation against a ground truth is possible. In [9] the authors propose to tag objects with sensors to automatically annotate when objects are manipulated (e.g. when the user picks up a cup to drink). This approach is limited to environments where all relevant objects can be tagged.

There is a need for real-time annotation solutions to reduce annotation effort and increase annotation quality.

III. EAR-WORN SENSING AND ANNOTATION

We used 3D CAD rapid prototyping to produce the ear-worn hardware shown in Fig. 1. Normal hearing people can use it to collect reference data relevant for the development of algorithms for multimodal HIs assuming that head movements do not differ between normal hearing and hearing impaired people.

A. Sensing and Data Handling

The device weighs 14.5 grams and its maximum dimensions are $53\,\text{mm} \times 30\,\text{mm} \times 10\,\text{mm}$, which is larger than a HI but similar to a common speech processor worn behind the ear by cochlear implant users. It is powered by a rechargeable Lithium-Polymer battery with a capacity of 110 mAh. The integrated miniaturized ETHOS IMU [10] comprises a low-power 16 bit dsPIC (MICROCHIP dsPIC33FJ128) with 16 kB on-board RAM and an integrated real time clock, a triaxial MEMS accelerometer (Linear Technology LIS3LV02DL) with a resolution of 16 bit, a triaxial magnetic field sensor (Honeywell HMC5843) with a resolution of 12 bit, a triaxial gyroscope (Invensense ITG-3200) with a resolution of 16 bit, a temperature sensor for automatic self-calibration, logic to write sensor data to a microSD-Card, an USB interface to download data in real-time or after the recordings, to configure the sampling rates of the sensors and to charge the battery, and an ANT+ module for low power wireless communication, e.g. to transmit sensor data to a mobile phone.

The system runtime for a battery capacity of 110 mAh is more than 16 hours for sampling 3D acceleration at a rate of 32 Hz. When all three sensors are sampled at a rate of 128 Hz the runtime is more than 5 hours.

B. Data Annotation: One-Button-Approach

We integrated a button into the device that the user can press once or multiple times to annotate the data in real-time (e.g. once for starting a certain activity; twice for stopping it). An advantage of the user annotating the data in real-time is high annotation quality compared to offline annotation approaches, where the user needs to remember his situation. The user interaction with the button is stored synchronized together with the sensor data. We opted for the one-button-approach to maintain a standalone system with an easy to use and as unobtrusive as possible annotation process. Alternatively, a smartphone application could be used to annotate data from our device at a finer granularity, but the resulting phone interaction can be considered as too obtrusive depending on the application. Figure 2 shows head movement data and annotation recorded with our device.

C. User Acceptance

We conducted a user study involving 21 normal hearing participants (17 male, 4 female, age 18–60) to evaluate the user acceptance of our system. We considered normal hearing participants in the study, because we assume that the acceptance is perceived similar to HI users. Normal hearing people use the system as well to achieve the necessary large quantity of reference data. In the first phase of the user study, the participants were asked to wear the device and perform everyday activities, paying special attention to comfort and hold of the device. The activities included walking (alone and while talking with a colleague), going up and down stairs, jumping and shaking the head. In a second

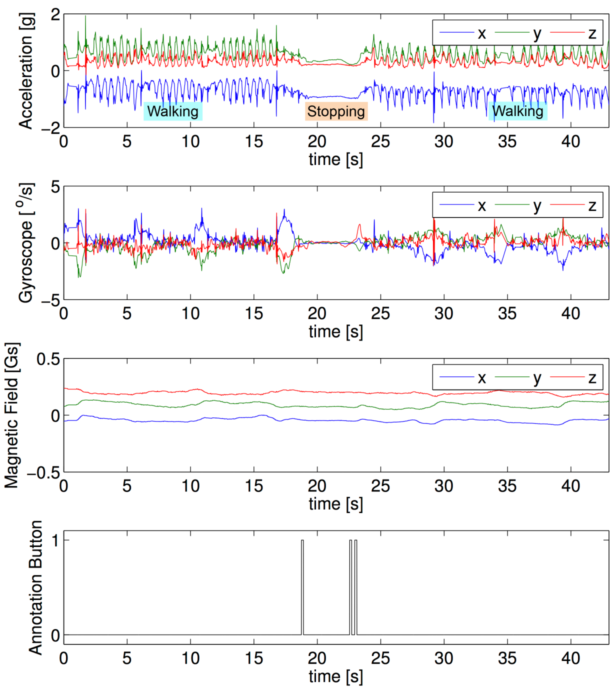

Fig. 2. Head movement data and data annotation by the user recorded with the ear-worn device.

phase, the participants were asked to evaluate the device in a questionnaire regarding its wearing comfort, noticeability, size, hold, unobtrusiveness and ease of use. For each topic they could choose their answers from four categories. We found that (number of occurrences in brackets):

- wearing the device was perceived as comfortable (14) or very comfortable (7)
- people did not find the device to be noticeable (9), or just very slightly (6)
- the device holds tightly and does not drop; even for strong shakes of the head, the hold was good (10) or very good (10) and we found the device to also hold tightly when going by bike or jogging over rough terrain
- the one-button approach for real-time data annotation is simple to learn and use (21); people believe that pressing the button behind the ear feels slightly eye-catching (12) or even striking (2) when moving in public
- about half of the participants (12) suggests that, while being already very comfortable to wear, the device may be further improved by shrinking its dimensions.

D. Limitations

The one-button approach for data annotation is simple to use but reaches its limits when a large amount of different classes needs to be annotated. However, in our applications it is sufficient, as only a small amount of classes will appear in the data sets to be recorded. If many annotation classes are required, the system can be used to set the class boundaries. In a consecutive offline step the already segmented annotation class instances can then be labeled manually, e.g. based on audio, video or experiment documentation notes.

One of the user's ears is occupied by the device, so only one additional regular HI can be worn. If two HIs are required for a given application, e.g., to record stereo sound, our system needs to be merged with the HI into a single device or in-the-ear HIs need to be used.

IV. UNSUPERVISED CLUSTERING OF HEAD MOVEMENT DATA

A. Data Set

We designed an experiment to investigate how movements of the HI user's head can characterize activities of daily living (ADL) that are relevant for the user's current hearing wish. The proof-of-concept experiment involved a single user with no hearing impairment and was tailored specifically to cover a wide range of ADL of elderly people since they form the largest group of HI users [11]. Over one hour of data has been recorded in real-life settings like restaurants, crowded pedestrian areas and tourist sites. Safety-critical situations when moving in the city traffic as well as rides by train and tram were included as well.

B. Method to Identify Clusters in Head Movement Data

In the analysis of the recorded data we focus on the head acceleration data (sampled at 128 Hz), because we found promising results with this kind of data with a wired system during a study in an office scenario [1].

Using a sliding window of 100 samples with a step size of 50 samples we calculated a set of 13 features from the head movement acceleration data comprising the mean, variance, mean crossing and zero crossing rate of the acceleration in x, y and z-direction, and the triaxial magnitude. We then applied Principal Component Analysis (PCA) [12] to the calculated features. PCA is a method to transform a high dimensional feature space into a lower-dimensional feature space to reveal structures contained in the higher-dimensional set. PCA identifies the transformed axis directions, called principal components (PCs), by means of singular value decompositions and ranks them according to the data's variances along their directions. Components that vary little in the transformed feature space (small singular values) are neglected to reduce the dimensionality, assuming that they do not add significant information to the PCs.

We plot the transformed feature space in a 3D coordinate system obtained by the three PCs with the highest singular values and color them according to their data annotation to identify clusters.

C. Results and Discussion

The result plots were obtained by plotting the coordinates of the first three PCs of the feature data. One data point represents a time interval of 781 ms. In addition, ellipsoids were fitted to visualize the distributions of the different data clouds and hence the clustering. The highest occurring singular values were 3.31, 1.83, and 1.51 and the corresponding PCs are composed mainly from the following features: mean-crossing x, mean y, mean z for PC1, mean x, mean-crossing y, zero-crossing y for PC2, and var x, zero-crossing x, mean-crossing y for PC3.

Fig. 3 shows the clustering of the user's modes of locomotion, i.e. walking, sitting and standing. The user's mode of locomotion is important for a new generation of HI applications as it can correlate with the user's hearing wish [2]: When the user is not moving, his hearing wish does statistically change less than during walking. By introducing corresponding priors the adaption of the HI can be tuned.

Fig. 4 shows the clustering of the user's hearing wish. We distinguish "noise comfort", which refers to a noisy environment where loud sounds need to be damped by the HI for user comfort, and "unintentional hearing", which refers to a situation where the wearer does not focus on a particular sound source and wishes to perceive sound as natural as possible. As "directed conversation" we define a conversation where the speakers are talking face-to-face.

The classes described above are reflected in clusters in the transformed feature space, which allows distinguishing them in later context recognition implementation. The overlap of the clusters will lead to a reduced recognition rate, depending on the classifier. Additional PCs and features can be taken into account to further reduce the overlapping regions of the clusters.

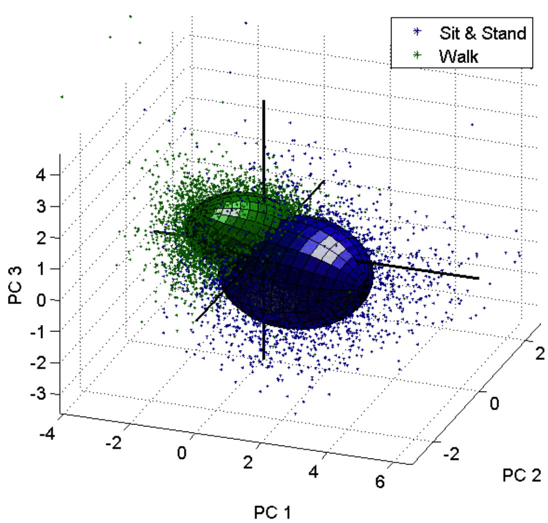

Fig. 3. Clusters in the head acceleration data (user's mode of locomotion)

V. CONCLUSION AND OUTLOOK

We presented a newly developed ear-worn sensing and annotation device, which is compact enough to be worn behind the user's ear. It allows us to unobtrusively capture head movements in real life situations in order to record reference data sets that are crucial for the development of algorithms for multimodal HIs. A user study with 21 participants showed that the device was perceived as comfortable and the one-button-approach for data annotation as easy to learn and intuitive. Unsupervised clustering of head acceleration showed a separation in the feature space of relevant user activities, which is a very promising result for the development of multimodal HIs. We believe the novel technology will help to push the boundaries of HI technology.

We plan to deploy our system to establish a large-scale database of reference data covering a variety of real life

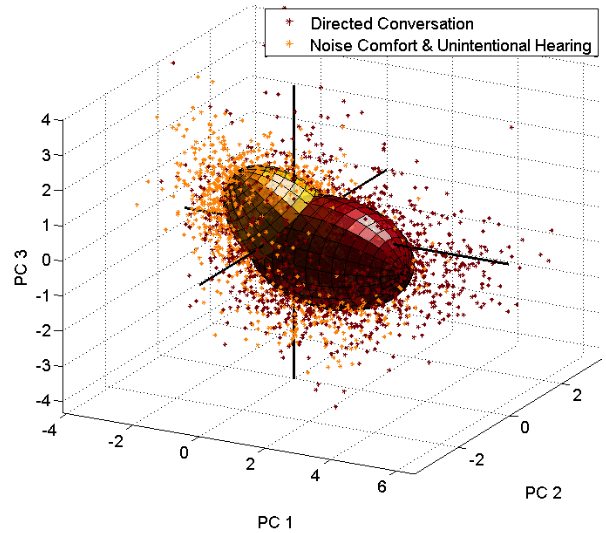

Fig. 4. Clusters in the head acceleration data (user's hearing wish)

settings to devise algorithms for automatic hearing program selection based on multimodal sensing. We further plan to investigate potential differences in the head movement of severe hearing impaired people and normal hearing people.

Besides the applications in the development of multimodal HIs the device can as well be used for other applications, e.g. fall and posture detection or movement analysis in sports.

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

REFERENCES

[1] B. Tessendorf et al, "Recognition of hearing needs from body and eye movements to improve hearing instruments," in *International Conference on Pervasive Computing*, 2011.
[2] ——, "Identification of relevant multimodal cues to enhance context-aware hearing instruments," in *6th International Conference on Body Area Networks*, 2011.
[3] L. Atallah, O. Aziz, B. Lo, and G.-Z. Yang, "Detecting walking gait impairment with an ear-worn sensor," *International Workshop on Wearable and Implantable Body Sensor Networks*, 2009.
[4] M. Poh, N. Swenson, and R. Picard, "Motion-tolerant magnetic earring sensor and wireless earpiece for wearable photoplethysmography," *Information Technology in Biomedicine*, 2010.
[5] E. Tamaki, T. Miyaki, and J. Rekimoto, "Brainy hand: an ear-worn hand gesture interaction device," in *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*. ACM, 2009, pp. 4255–4260.
[6] J. Patterson, D. McIlwraith, and G. Yang, "A flexible, low noise reflective ppg sensor platform for ear-worn heart rate monitoring," in *Wearable and Implantable Body Sensor Networks*. IEEE, 2009.
[7] D. Roggen et al, "Collecting complex activity data sets in highly rich networked sensor environments," in *7th Int. Conf. on Networked Sensing Systems*, 2010.
[8] S. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1345–1359, 2009.
[9] A. Calatroni et. al., "A methodology to use unknown new sensors for activity recognition by leveraging sporadic interactions with primitive sensors and behavioral assumptions," in *Opportunistic Ubiquitous Systems Workshop*, 2010.
[10] H. Harms et al, "ETHOS: Miniature Orientation Sensor for Wearable Human Motion Analysis," in *IEEE Sensors*, 2010.
[11] S. Kochkin, "MarkeTrak VIII: 25-year trends in the hearing health market," *Hearing Review*, vol. 16, no. 10, 2009.
[12] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, 1987.