Audio Event Detection for In-Home Care

P.W.J. van Hengel[1], J. Anemüller[2]

[1] *Fraunhofer Institut Digitale Medientechnologie, Project group Hearing, Speech and Audio Technology, 26129 Oldenburg , Germany, Email: peter.vanhengel@idmt.fraunhofer.de*

[2] *Section Medical Physics, Carl-von-Ossietzky-university, 26111 Oldenburg, Germany, Email: joern.anemueller@uni-oldenburg.de*

Introduction

One in eight persons attending hospital following an accident in the home, are aged 65 and over. This number will increase with the demographic change and the desire of the elderly to remain in their own homes. Hence, the demand for unobtrusive monitoring systems based on autonomous sensor technologies will increase. Such systems can limit observation by a human operator to cases when there is evidence of an incident, by responding to events in a 'human-like' fashion. To do so, the system must pick up situations that would also attract human attention, including situations it has not encountered before.

Hearing is an important sense in terms of steering the attention of a human observer. For this reason the detection of sound events that could indicate an incident is a key part of such a monitoring system. An acoustic system designed for the detection of verbal aggression has shown good results. Now a system is developed for in-home monitoring, in which event detection is combined with acoustic localization and integrated with video analysis and artificial intelligence. The resulting methods will be tested in the context of in-home care under real-life conditions. For this purpose a realistic test-home has been constructed.

The demographic change

In the next decennia Europe is facing an increase in the average age of the population. Especially the group of the 'very old' – over 80 – in Germany has increased more than six-fold between 1910 and 2000. Model calculations show that in 2030 approximately 50% of the German population will be over 47 years of age. The "Alterslastquotient", which is the ratio of the number of people over 64 and the number of people between the ages of 18 and 64 – the professional population –, will increase from 25% to almost 44% in only three decades.

Old age for many people is unfortunately accompanied by a loss of physical en psychological abilities. Especially in the group of the 'very old' we see such afflictions as cardiovascular problems, diabetes, dementia, hearing loss etc usually occurring in combination. This implies that the number of people in need of some form of care will increase by 58% between 2005 and 2030.

On the other hand, most elderly people are living in their own private homes, and 73% wants to continue to do so, even if their physical or psychological condition deteriorates. [1]. Moving to a care institution is associated with a loss of autonomy and a reduction of the quality of life. [2]

The own home becomes more and more the central domain in the lives of the elderly. The reduced mobility due to physical limitations is aggravated by a fear to leave the familiar surroundings. Although almost one fifth of elderly people think of technology as threatening [3], it is clear that the application of modern technologies for monitoring is required.

Unobtrusive monitoring

The term Ambient Assisted Living (AAL) was coined to describe the specific sub-domain of ambient intelligence devoted to developing technologies for the design of environments for the aging. Amongst the goals of AAL are:

- increasing the comfort of living (e.g. by automatic control of room lighting);
- increasing security (e.g. by detecting water leakage)
- supporting health care (e.g. by monitoring vital signs and by detecting medical emergencies).

For the first two of these goals many technical solutions already exist and the major challenge is to create interoperability and ease-of-use. For the last domain many barriers still prevent solutions to be implemented on a large scale. These barriers range from a lack of good business models to technical problems and issues with privacy and (legal) responsibility.

Especially when it comes to monitoring systems the aspects of privacy and user acceptance are critical. A study that is currently being conducted on the acceptance of camera surveillance systems and body-worn sensors shows that as age increases and the perceived health is deteriorating, such devices are more accepted by elderly [4]. However, in combination with the difficulties of state-of-the-art video analysis systems to detect incident situations it is clear that video monitoring cannot provide an acceptable system for the detection of emergency situations. The use of other sensor data to infer e.g. (in)activity patterns, although promising, is incapable of responding quickly and reliably to such emergencies.

The challenge for AAL is therefore mostly the detection of (medical) emergencies using a system which is truly ambient (that is, does not impede the user in any way), does not invade privacy any more than absolutely necessary and is reliable. As some of the elderly refer to it: they would like to have 'someone within earshot'.

Audio Incident detection

It appears, to anyone not involved in sound analysis or AAL, to be perfectly natural to use sound as a first indicator of incidents. After all, for human observers it is the sound that usually attracts our attention.

Based on this notion the development of methods for the detection of incidents, based on acoustic information, was started in 2003 with a short study on the human capabilities to detect aggression in voices [5]. An aggression detection system based on these methods was evaluated in the city centre of Groningen in 2006, showing the potential of using acoustic detection as a first-warning system for video surveillance [6].

Other systems for the detection of sound sources have been developed, based on neural network theory, and have been shown to be effective. However, for the detection of incidents a new approach was chosen. The main problem with the use of neural-network based approaches is their dependence on large sets of training data. In the case of incidents these are difficult to obtain, unless one resorts to staging and acting. However, for human listeners – in the case of aggression at least – it is not difficult to hear the difference between acted and real scenes, indicating there are important acoustic differences. Also, it is known that neural network based approaches lack the robustness required to use the system in various acoustic environments.
For this reason an approach based on human auditory processing and human auditory attention was chosen. Using a model of the human inner ear and concepts of auditory grouping so-called auditory events can be detected. These events are what humans perceive as individual sounds (e.g. the sound of a passing car, a car horn, a word, a glass shattering on the ground, etc)
Using models of human perception several psychophysical quantities can be derived describing these events. These psychophysical quantities can be combined to give an attention value to the event. And based on the attention value the system can decide whether to further analyze the sound, or to send an alert or alarm.
The shortcoming of neural network based technologies can be solved in a different way. Instead of having the system search for a predefined type of sound indicative of an incident, the system could be trained to give an alert or alarm when it encounters an unknown sound. This approach sounds simple enough and seems to correspond well to the natural manner in which biological systems deal with the variety they encounter in everyday life. However, on closer examination it turns out to be very difficult to ascertain whether an event can not be recognized. Simply using a goodness-of-fit for the detector does not work since there can be a myriad of reasons for a low goodness-of-fit, especially in realistic circumstances.
An interesting new approach to this 'reverse' type of incident detection is taken in the DIRAC project [7], where recognition systems at different levels of complexity are combined to detect incongruencies. A good response of general models trained on individual building blocks, but a failure of models trying to find the expected combination of these building blocks, is a clear indication that something is wrong. This approach has the potential to be used at various levels of complexity, ranging from the detection of an unexpected sound to the detection of changes in behavioral patterns happening over a period of many weeks.

The IDEAAL room

In order to test new technological applications for AAL, such as an acoustic incident detection system, a special 'room' was constructed in the Lower Saxony Research Network Design of Environments for Ageing [8]. Apart from offering the scientist every opportunity to incorporate their hardware and test various methods of monitoring and interacting with the elderly, the 'room' is actually a fully functioning and self-contained apartment.

Figure 1: Living room of the IDEAAL appartment

Figure 2: Fully functional kitchen of the IDEAAL appartment

Scenarios based on studies of the most common problems faced by the elderly in their home environment, user studies and expert interviews, form the basis for the deployment of technologies. The effectiveness and implications of these technologies will be tested in an environment that can hardly be more realistic.

In the living room and kitchen, microphones will be installed to monitor the activity of the elderly occupant and signal potential incidents. In addition the possibility to use these microphones for a speech interface with the system components requiring user-system-interaction will be tested. Also any acoustic signals indicative of the health status will be recorded, e.g. coughing or sneezing.

References

[1] Grauel, J. and Spellerberger, A. (2007): Akzeptanz neuer Wohntechniken für ein selbständiges Leben im Alter – Erklärung anhand soziostruktureller Merkmale, Technikkompetenz und Technikeinstellung. In: Zeitschrift für Sozialreform Jg 53 H 2, 191-215.

[2] Schneekloth, U. and Wahl, H-W. (2008): Selbständigkeit und Hilfsbedarf bei älteren Menschen in Privathaushalten, Pflegearrangements, Demenz, Versorgungsangebote. Stuttgart: Kohlhammer.

[3] Friesdorf, W and Heine, A. (2007): 'sentha' – seniorengerechte Technik im häuslichen Alltag. Ein Forschungsbericht im integrierten Roman. Springer Verlag New York

[4] Meis, M., Gövercin, M. & Költzsch, Y. (2009): Anforderungsanalyse AP 4 zur Sturzerkennung und –prädiktion. Interner Forschungsbericht (unpublished).

[5] Huisman, M. (2004): Akoestische Effecten van Emoties in Spraak: De Waarneming van Verbale Agressie (*Acoustic effects of Emotions in Speech: the Perception of Verbal Aggression*). master thesis Department of AI, University of Groningen

[6] van Hengel, P. and Andringa, T. (2007): Verbal aggression detection in complex social environments. 2007 IEEE International Conference on Advanced Video and Signal based Surveillance.

[7] http://www.diracproject.org

[8] http://www.altersgerechte-lebenswelten.de