Acoustical Method for Objective Food Intake Monitoring Using a Wearable Sensor System

Sebastian Päßler, Wolf-Joachim Fischer
Fraunhofer Institute for Photonic Microsystems
IPMS
Dresden, Germany
Email: Sebastian.Paessler@ipms.fraunhofer.de

Abstract— **A method for non-invasive monitoring of human food intake behavior and long-term dietary protocol has been developed by the sole use of chewing and swallowing sound sensors. A hierarchical model of daily food intake is introduced for dietary protocol. Microphones were found to be more useable for this recording task than accelerometers. Hence, a novel sensor system has been built containing two microphones in a hearing aid case in order to record chewing and swallowing sounds in the ear canal and environmental noise, respectively. Records of the intake sounds of 51 participants consuming seven types of food and one drink were taken for the development of detection and classification algorithms.**

Keywords- Chewing sound, dietary protocol, food intake model

I. INTRODUCTION

According to the World Health Organization (WHO), the number of nutrition-related diseases like obesity and Diabetes mellitus raised significantly in the last centuries [1]. Therefore, nutrition monitoring gains strong significance. In 2005 approximately 1.6 billion adult people worldwide were overweight and at least 400 million of them reached the state of obesity. Projections of up to 2.7 billion overweight adults and 700 obese people in 2015 are made by the WHO. This trend is said to be a result of increasing energy intake and decreasing energy consumption by physical activity. A precise timesaving method to protocol nutritional intake is needed. In contrast, today's method for monitoring food consumption is done manually by information sampling on questionnaires. This very time-consuming process causes tiredness and often denial in patients. Furthermore, accuracy of this method often is low and depends to the effort of patients. Underestimation of the amount of food consumption of up to 50 % is not unusual [2]. Things get worse in ambient assisted living or retirement homes, where nursing auxiliaries have to do protocol for elderly people. Often patients are asked only once a day and thus snacks are forgotten in records and the amount of consumed food could only be guessed. A device that records and protocols information about the moments of eating as well as type and amount of consumed food automatically would overcome these disadvantages.

This paper is organized as follows: Section 2 presents current studies in the field of food intake recognition. A hierarchical model of daily food intake is presented in Section 3. In Section 4, the sensor election process is shown. Recording methods are presented in Section 5, followed by the Conclusion.

II. RELATED WORK

Different authors have already paid attention to the problem: While food identification with photos taken by PDAs [3] or use of RFID technology [4] means higher effort for users or food manufacturers, an analysis of the sounds generated in the process of chewing would overcome these disadvantages. Drake was the first who investigated acoustics of chewing hard and crisp food [5]. He showed that chewing of each food type generates characteristic sounds. Amft et al. used this approach for recognition of food consumption in investigations with microphones integrated in headphone cases [6]. Sounds were recorded in the outer ear canal and swallowing signals were taken from an EMG sensor at the throat. Shuzo et al. worked with the approach of acoustical classification of chewing sounds [7]. They used a bone conduction microphone in the ear canal to detect moments of eating, to count the number of chewing strokes and to distinguish between two different food categories. Participants should consume a hard and a soft food, drink a liquid and read. Lopez-Meyer et al. used microphones to record chewing and swallowing sounds and detect periods of food intake by evaluation of the instantaneous swallowing frequency [8]. Microphones used here are in-ear microphones and throat microphones. Additionally, strain sensors on the junction of temporal bone and mandible are used to record chewing movement.

III. FOOD INTAKE MODEL

The procedure of food intake is related to two sub-processes, which are essential for it: chewing and swallowing. Chewing is defined as the crushing and size reduction of larger pieces of solid food and their mixture with saliva [9]. Swallowing consists of forming a food bolus and its transport into the esophagus. Food crushing of solid food is caused by pressure and shear motion of the molars, that affect at the upper and lower side of the food piece in the mouth. Caused by this, cracks in the food arise and propagate. Finally, the piece is disrupted or broken. Because of this, teeth do not move continuously towards each other and receive little vibrations by every

This work was supported by the DFG research training group "Nano- and Biotechniques for Electronic System Packaging" (DFG 1401/1).

PervasiveHealth 2011, May 23-26, Dublin, Republic of Ireland

DOI 10.4108/icst.pervasivehealth.2011.246029

single rupture or breaking process in the piece of food. The vibrations are propagated in the skull bone.

In [8], Amft introduced a model of the chronological sequence and hierarchy of food intake. He termed the smallest unit of a record of chewing sounds, which could be separated as a single event, a *chew event* or a *swallowing event*, depending on their belonging to the sub-processes of food intake. Additionally, he introduced the *intake event* as a gesture necessary to carry a piece of food or a drink into the mouth. The process consisting of the chronological sequence of an intake event, chew events and one or several swallowing events for solid food or the same sequence without chew events for liquid food and drinks, respectively, is referred to as *intake cycle*. The model depicts, that the type of consumed food can not change during an intake cycle. Hence all chew and swallowing events of one intake cycle refer to the same type of food or drink, respectively. Referring to Amft, a chew event consists of four phases, which can be described as the recurring sequence of

- closure of the jaw, whereby food is crushed,
- a pause without movement of the jaw, while teeth are pressed together,
- opening of the jaw and
- another pause, while the opened jaw is in rest.

Usually the strongest microphone signal is achieved while closing the jaw, because food is crushed and typical chewing sounds are generated in this process. In both pause phases no sounds related to food intake occur. Amft isolates single chew events in the pause phase while the jaw is opened.

This introduced model is the base of the food intake model in this work, which also depends on chew and swallowing events. But a few changes are made to it as described below:

(1) Intake events are not part of the model in this work.

(2) Additional, higher hierarchical levels above the intake cycle are introduced, according to the aim of dietary protocol. These are the levels of “meal” and of “daily protocol”.

(3) The chew events of one mastication cycle are differentiated further depending on their sound properties caused by the progress of food crushing.

Change (1) results from the fact, that no movement sensors are used to record and recognize the intake gestures.

The extensions of new hierarchical levels mentioned in change (2) are shown in Fig. 1. Several chronologically concentrated chewing and swallowing cycles, which occur in the same period of time, belong to similar food. They could be divided by short periods of speech or environmental sound, are modeled as “meal”. Depending on the amount of consumed food and the point of time of its occurrence, the meal could be marked as principal meal (i.e. breakfast, lunch, dinner) or a short snack. These times of concentrated food or liquid intake could be summarized in a “daily protocol” of the food intake during 24 hours of one day. “Daily protocol” is the individual protocol of different meals at several discrete points of time in a whole day’s food intake record. This is the abstraction level of standard intake questionnaires.

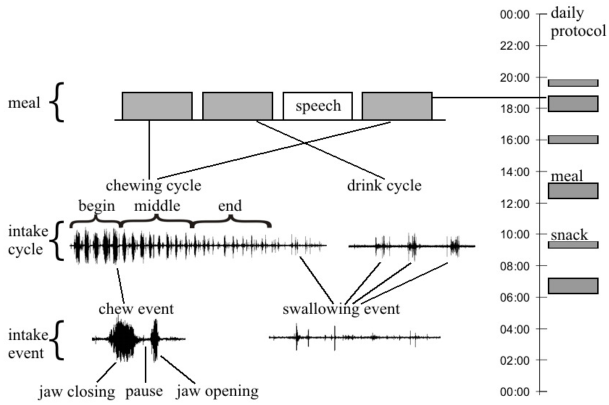

Figure 1. Hierarchical levels of the automatic intake protocol

It is mentioned in change (3) that the properties of the chewing sounds of one type of food change during the mastication cycle caused by the progress of food crushing and the mixture with saliva. This is also observed in [10]. Lee et al. investigated the chewing sounds recorded outside of the closed mouth of participants, who were chewing potato chips and tortilla chips. They discovered that the signal energy of the spectrum of a chew event was reduced in the progress of food crushing. By this, signal energy in higher frequencies fades away and the spectral composition of the spectrum moves towards lower frequencies. This change is expressed in the sound of the record in the way that the sound of the first chews of a mastication cycle can be described as “dry, hard scrunching” while the sound of the following chews is described as getting “wetter” and more “smacking”. It is caused by the mixture of the food particles with saliva in the mouth. In our work, this change of the spectral composition of the chewing sounds during the mastication cycle is modeled by the introduction of three phases of different chewing sounds in one mastication cycle:

- *begin*: food is mainly compact and has to be crushed; chewing sounds can be described as „dry, hard scrunching“ and their signals have significant signal energy in higher frequency ranges
- *middle*: advanced food crushing, mixture with saliva starts; chewing sounds are dampened and their signals have lower signal energy in the higher frequency range
- *end*: food is papescent and mixture with saliva continues; chewing sounds can be described as “wet, smacking” and their signal energy in higher frequency range have almost vanished

In this model swallowing events can occur either in the phase *middle* or in *end*. Usually, a mastication cycle is finished with a swallowing event.

IV. SENSOR ELECTION

It is the aim of this project to deliver a proved conception for a portable device, which could detect food intake of a patient automatically and extract parameters of strong interest for

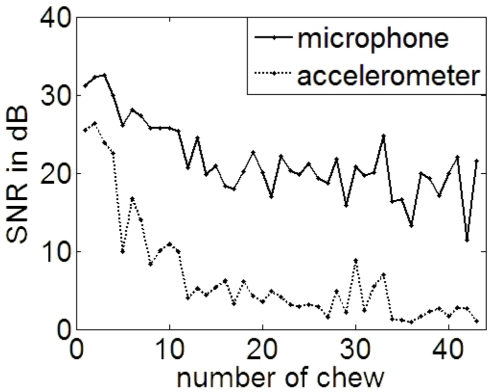

Figure 4. A-posteriori SNR when participant ate a piece of carrot

intake monitoring from the recorded signals. Such parameters are the moment of food intake, type and amount of the consumed food and the instantaneous chewing frequency. A non-invasive, unobtrusive and low disturbing device could raise the patient's acceptance to wear it. Automatic recording and signal processing of the recorded data with high reliability and accuracy are aspired.

In our studies, we mainly investigated two promising sensor concepts to record and analyze human chewing: Chewing and swallowing sounds can be recorded using in-ear microphones or analyzed using recordings of the transmitted vibrations in the skull from accelerometers applied at the human head. Both sensors have the advantages to be non-invasive and to have the potential of miniaturization. Accelerometers cover the additional advantage of nearly free choice of a useable application place on the human head. This choice could be made taking account of high signal quality and high acceptance of the patient. Literature indicate that microphones only record strong chewing sound signals compared to environmental sounds if the microphones are applied to the outer ear canal [11].

To choose a practical sensor we investigated some small sensor types with the potential of integration in a portable sensor system. For this, only sensors with their highest dimension not larger than 1 cm are used. The miniaturized electret microphone FG-23329-CO5 from Knowles Acoustics and the accelerometer PKGS-90LD-R from Murata Manufacturing Co., Ltd. were chosen and applied to an analog preamplifier and filter. The amplified and filtered signals were recorded simultaneously with the standard sound card of a notebook computer. Application places under investigation were the outer ear canal of the left ear for the microphone and on the skin covering the mastoid part of the left hand temporal bone (in occipital direction of the outer ear) for the accelerometer. Signal quality and amplitude of chewing and environmental sounds were compared qualitatively and the a-posteriori SNR was computed for every chew event signal. Signal amplitudes of the first few chews of hard texture food like carrots were well noticeable in the waveform diagram and both signals' quality was comparable according to their tone. But after the first six to ten chews, signal amplitude of the accelerometer's record decreased rapidly and could hardly be distinguished from noise. When the participant ate food with a soft texture like chocolate or a banana, the signal also almost disappeared in the noise. The a-posteriori SNR of the microphone's signal was on average 10 to 20 decibels higher than the SNR of the accelerometer's signal at the same

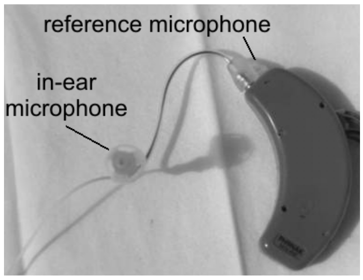

Figure 3. Hearing aid package with both integrated microphones

chew. The SNR of the microphone's signal ranges from 10 to 25 decibels, whereas the SNR of the accelerometer's signal ranged from 5 to 10 decibels. Fig. 2 shows the SNR of the microphone's signal and the accelerometer's signal over the number of the chew events in a mastication cycle when the participant ate a piece of carrot. In contrast to the microphone's signal, neither swallowing sounds nor sounds caused by the environment or the participant's speech could be noticed with strong amplitudes in the accelerometer's signal. While this fact is a great advantage of the accelerometer, the swallowing sounds cannot be disclaimed in the evaluation of food intake. Mainly due to these disadvantages of the accelerometer in matters of the signal quality we decided to use a microphone in our device for intake monitoring.

In respect of the aspired usage of the new device at patients in ambient assisted living and retirement homes, we decided to take a design, which is common under these circumstances and hence less flashy than other designs. Miniaturized electret microphones were integrated into a standard hearing aid case (Phonak SuperFont), because hearing aids are quite commonly used by elderly people. The earpiece of the hearing aid could be applied into the bony part of the ear canal to record the sound emitted at this place. Additionally, in hearing aid packages there is an opening for the hearing aid microphone. This opening can be used to integrate a second microphone in our design to record sounds from the environment synchronously to the sound in the ear canal. With this system, it is expected to ease the decision, if a sound was caused by the patient's body or by the environment. The in-ear microphone at the top of the earpiece of the hearing aid is covered using a standard silicon soft tip for hearing aids to protect it from ear wax. Fig. 3 shows a photo of the hearing aid package with both integrated microphones.

Fig. 4 shows a block diagram of the sensing path for time-synchronous records of the in-ear microphone and the reference microphone in the hearing aid package. Sampling was done using both channels of the standard sound card of a notebook

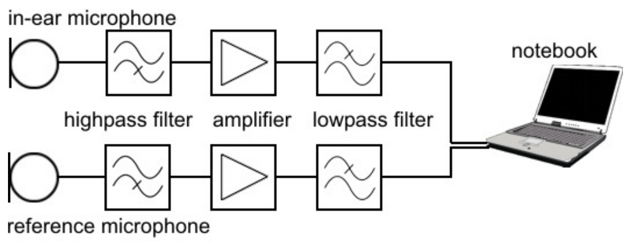

Figure 2. Block diagram of the sensing path for time-synchronous microphone records

computer. Sampling rate is 11,025 Hz and the resolution is 16 bit.

V. RECORDING METHODS

To build a database for the implementation and evaluation of detection and classification algorithms, records of chewing and swallowing sounds of various participants are carried out using fixed instructions. The sensor system introduced in Section 4 was used. The in-ear microphone was applied to the participant's outer ear canal of the right ear while the hearing aid package with the reference microphone was applied behind the outer ear. 51 participants aged 15 to 77 years (mean: 34.8) are instructed to eat ten pieces of seven types of food (potato chips, peanut, walnut, carrot, apple, chocolate, pudding), respectively, and to drink 30 sips of a drink of their choice (juice or water). Food types were excluded, if participants expressed a strong dislike or were allergic of them. All participants had natural dentition. Records were done in a quiet recording room or office room. Participants should sit at a table and eat the food in single pieces. They were told to chew and swallow as they usually do. Chewing and swallowing sounds at the in-ear microphone and environmental sounds at the reference microphone were recorded synchronously. Participants were told not to talk during the record and to avoid any kind of disturbing sound. Overall, 21 hours of sound data were recorded. All records were annotated manually using the model presented in Section 3. Chew events were grouped to the single consumed foods and the phases *begin*, *middle* and *end* of the mastication cycle, respectively, depending on their sound and short time spectrum. Swallowing sounds were labeled as *swallow* and not differentiated further. Silent periods without significant sounds between the single chew or swallowing events were labeled as *zero* and sounds with no relation to the chewing or swallowing process (e.g. environmental sound, speech, noise) were labeled as *garbage*. The database contains overall 71168 chew events, 7087 swallowing events and 3967 events labeled as *garbage*. Future research on automatic detection and recognition algorithms will use this data and the manual annotation as base for model training and signal recognition, respectively.

VI. CONCLUSION

A method for non-invasive monitoring of human food intake behavior and long-term dietary protocol is necessary especially under the conditions of ambient assisted living or retirement homes. The evaluation of recorded chewing and swallowing sounds is expected to be a promising approach towards this goal. A hierarchical model of daily food intake suitable for automatic intake monitoring is introduced. We investigated a microphone and an accelerometer for their feasibility and found the microphone to be the best. A sensor system was built using two microphones for chewing and swallowing sounds, and for environmental sounds, respectively, integrated into a hearing aid case. A database of recorded and annotated mastication cycles has been created. 51 participants were advised to eat seven types of food and consume one drink, while records were taken from the mastication and swallowing sounds. The annotation algorithm for the records of the single intake cycles is introduced. The chew events were labeled based on both the type of food eaten and the phase in the mastication cycle. Three phases were introduced in this way. These data will be the basis for further studies on automatic evaluation of the human intake behavior.

ACKNOWLEDGMENT

The authors would like to express their thanks to Matthias Wolff from the Institute of Acoustics and Speech Communication at Technische Universität Dresden for his support with acoustic and signal processing technologies.

REFERENCES

[1] World Health Organization, "Obesity and overweight: What are overweight and obesity?", *Fact sheet N°311*, Geneva: WHO, September 2006.

[2] D. A. Schoeller, "Limitations in the assessment of dietary energy intake by self-report", *Metabolism*, Vol. 44, pp. 18–22, February 1995.

[3] J. Beasley, "The pros and cons of using pdas for dietary self-monitoring.", *J Am Diet Assoc*, Vol. 107, No. 5, p. 739, 2007.

[4] K.-H. Chang, S.-Y. Liu, H.-H. Chu, J. Y. Hsu, C. Chen, T.-Y. Lin, C.-Y. Chen, P. Huang, "The diet-aware dining table: Observing dietary behaviors over a tabletop surface." In K. Fishkin, B. Schiele, P. Nixon, and A. Quigley, ed., *PERVASIVE 2006: Proceedings of the 4th International Conference on Pervasive Computing*, vol. 3968 of *Lecture Notes in Computer Science*, pp. 366–382. Springer Berlin, Heidelberg, May 2006.

[5] B. Drake, "Food crushing sounds. an introductory study", *J Food Sci*, vol. 28, no. 2, pp. 233–241, 1963.

[6] O. Amft, "Automatic dietary monitoring using on-body sensors - Detection of eating and drinking behaviour in healthy individuals", Ph.D. Thesis, ETH Zürich, 2008.

[7] M. Shuzo, S. Komori, T. Takashima, G. Lopez, S. Tatsuta, S. Yanagimoto, S. Warisawa, J.-J. Delaunay and I. Yamada, "Wearable Eating Habit Sensing Using Sound Information", in Proc. of *the 2009 Conference on Micromechatronics for Information and Precision Equipment* (MIPE 2009), Tsukuba International Congress Center, Ibaraki, Japan, pp. 221-222, June 17-20 2009.

[8] P. Lopez-Meyer, S. Schuckers, O. Makeyev and E. Sazonov, "Detection of Periods of Food Intake using Support Vector Machines", in *32nd Annual International Conference of the IEEE EMBS*, Buenos Aires, Argentinia, pp. 1004-1007, August 31 – September 4 2010.

[9] P.W. Lucas, J.F. Prinz, K.R. Agrawal, and I.C. Bruce, "Food physics and oral physiology", in *Food Quality and Preference* vol. 13, issue 4, pp. 203-213, June 2002.

[10] W.E. Lee III, A.E. Deibel, C.T. Glembin and E.G. Munday, "Analysis of Food Crushing Sound During Mastication: Frequency-Time Studies", in *Journal of Texture Studies*, vol. 19, issue 1, pp. 27-38, January 1988.

[11] O. Amft, M. Stänger, P. Lukowicz and G. Tröster, „Analysis of Chewing Sound for Dietary Monitoring", in *UbiComp 2005: Proceedings of the 7th International Conference on Ubiquitous Computing*, pp. 56-72, Springer Berlin, Heidelberg, September 2005.