Analysis of Chewing Sounds for Dietary Monitoring

Oliver Amft[1], Mathias Stäger[1], Paul Lukowicz[2], and Gerhard Tröster[1]

[1] Wearable Computing Lab.,
Swiss Federal Institute of Technology (ETH) Zürich, Switzerland,
http://www.wearable.ethz.ch
{amft,staeger,troester}@ife.ee.ethz.ch
[2] Institute for Computer Systems and Networks, University for Health Sciences,
Medical Informatics and Technology (UMIT), Hall in Tirol, Austria,
http://csn.umit.at
paul.lukowicz@umit.at

Abstract. The paper reports the results of the first stage of our work on an automatic dietary monitoring system. The work is part of a large European project on using ubiquitous systems to support healthy lifestyle and cardiovascular disease prevention. We demonstrate that sound from the user's mouth can be used to detect that he/she is eating. The paper also shows how different kinds of food can be recognized by analyzing chewing sounds. The sounds are acquired with a microphone located inside the ear canal. This is an unobtrusive location widely accepted in other applications (hearing aids, headsets). To validate our method we present experimental results containing 3500 seconds of chewing data from four subjects on four different food types typically found in a meal. Up to 99% accuracy is achieved on eating recognition and between 80% to 100% on food type classification.

1 Introduction

Healthy lifestyle and disease prevention are a major concern for large portions of the population. Considering the worrying trend of sky-rocketing health care costs and the ageing population, these are not just personal but also important socio-economic issues. As a consequence all concerned parties: individuals, health insurance and governments are willing to spend considerable resources on tools that help people develop and maintain healthy habits. In Europe a considerable portion of research funding in this area is directed at mobile and ubiquitous computing technology. Within this program our group is involved in the 34 Million Euro MyHeart project that includes 35 medical, design, textile and electronics related research institutions and companies.

The aim of the consortium is to develop schemes that combine long term physiological monitoring and behavioral analysis with a personalized direct or professional-observed feedback to help users reduce their risk of cardiovascular disease. As is well known, the three main aspects that need to be addressed are

M. Beigl et al. (Eds.): UbiComp 2005, LNCS 3660, pp. 56–72, 2005.

stress, exercise and diet. In the project our group focuses on the later. Our aim is to develop wearable sensing technology to aid the user in monitoring his eating habits. In this paper we report on results of the first stage of this work: using wearable microphones to detect and classify chewing sounds (called mastication sounds) from the user's mouth.

1.1 Dietary Monitoring

Dietary monitoring includes a variety of factors starting from the diet composition to frequency, duration and speed of eating, all of which can be relevant health issues. Today such monitoring is almost entirely done 'manually' by user questionnaires. Electronic devices are at best used as intelligent log books that can derive long term trends, calculate calories from entered data and give simple user recommendations. The collection and entry of the data has to be done by the user which involves considerable effort. As a consequence, as anyone who has ever attempted a diet knows, compliance tends to be very poor.

Since prevention involves the adaptation of a healthier lifestyle, long term, quasi permanent monitoring (months or years) is needed to really make an impact on the risk of cardiovascular diseases. Thus any, even very rudimentary, tool that reduce the effort and interaction involved in data collection and entry could make a big difference.

1.2 Automating Dietary Monitoring

The ultimate goal of a system that precisely and 100% reliably determines the type and amount of all and any food that the user has consumed is certainly more of a dream then a realistic concept. However, we believe that with a combination of wearable sensors and a degree of environmental augmentation useful assistive systems are conceivable. On one hand, such systems could provide a rough estimate on the food consumption much like many today's physical activity monitoring devices provide only a rough guess of the caloric expenditure. On the other hand, it could be used as an entry assistant that, at the end of the day, would present the user with its best guess of when, how much, and what he has eaten and ask him to correct the errors and fill the gaps.

Overall we imagine such a non-invasive dietary monitoring support system to rely on the following three components:

1. Monitoring of food intake through appropriate wearable sensors. The main possibilities are
   (a) detecting and analyzing chewing sounds,
   (b) using electrodes mounted on the base of the neck (e.g in a collar) to detect and analyze bolus swallowing,
   (c) using motion sensors on hands to detect food intake related motions.
2. Monitoring food preparation/purchase through appropriate environmental augmentation. Here, approaches such as using RFID-tags to recognize food components or communicating with the restaurant computer to get a description and nutrition facts of the order are conceivable.

3. Including user habits and high level context detection as additional information sources. Here, one could accentuate the fact that eating habits tend to be associated with locations, times and other activities. Thus information on location (e.g in the dining room sitting at the table), time of day, other activity (unlikely to eat while jogging) etc. provide useful hints.

1.3 Paper Contributions

In the paper we concentrate on the first component of the envisioned system: food intake detection. Specifically, we consider the detection and classification of chewing sounds. To this end the paper presents the following results:

1. We show that good quality chewing sound signal can be obtained from a microphone placed in the ear canal. Since much of the acoustic signal generated by mechanical interaction of teeth and food during occlusion is transmitted by bone conduction, these sounds are actually much stronger than the speech signal. At the same time the location is unobtrusive and proven acceptable in applications such as hearing aids or recent high end mobile phone headsets.
2. We show that chewing sequences can be discriminated from a signal containing a mixture of speech, silence and chewing.
3. We present a method that detects the beginning of single chews in a chewing sequence.
4. We show that chewing sound based discrimination between different kinds of food is possible with a high accuracy.

For the above methods we present an experimental evaluation with a set of four different food products selected to represent different categories of food that might be present in a meal. The experiments consists of a total of 650 chewing sequences, from 4 subjects that amount to a total of 3500 seconds of labeled data. We show that recognition rates of up to 99% can be achieved for the chewing segment identification and of between 80 and 100% for the food recognition.

Overall, while much still remains to be done, our work proves the feasibility of using chewing sound analysis as an important component in a diet monitoring system. An important aspect of our contribution is the fact that the type of information derived by our system (what has actually been eaten) is very difficult to derive using other means.

1.4 Related Work

Activities of daily living are of central interest for high-level context-aware computing. Information acquisition can be realized by distributing sensors in the environment and on the human body. Realization of intelligent environments have been studied, e.g. in the context of smart homes [1] and mobile devices [2]. These works are generally focused on enhancing the quality of life, e.g. for independent living [3,4]. Smart identification systems have also been developed [5] which may provide information associated to nutrition phases, e.g. smart cups [6].

The interaction of chewing, acoustic sensation and perception of textures in food has been studied intensively in food science. Work in this area has been dedicated mainly to the relation of chewing sounds on the sensation of crispness and crunchiness. This was done by investigating air-conducted noises produced during chewing [7, 8] or by instrumental monitoring of the deformation under force [9,10,11,12] and studying correlation with sensory perception [13,14]. The loudness of a foodstuff during deformation depends mainly on the inner structure, i.e. cell arrangement, impurities and existing cracks [15]. Wet cellular materials, e.g. apples and lettuce, are termed wet crisp since the cell structures contain fluids whereas dry crisp products, e.g. potato chips have air inclusions [16]. A general force deflection model has been proposed [17] interpreting the acoustic emissions as micro-events of fracture in brittle materials under compression.

Initially Drake [9] studied the chewing sound signal in humans when chewing crisp and hard food products. It was found that a normal chewing cycle after bringing the food piece to the mouth cavity can be partitioned into two adjacent phases: Gross cutting the ingested material and conversion in fine grained particles. This process is understood as a gradually decomposition of the material structure during chewing and is audible as a decline of the sound level [9]. A swallowable bolus is formed after a certain level of lubrication and particle size has been reached. A first attempt was made by DeBelie [18] to discriminate two classes of crispness in apples by analyzing principal components in the sound spectrum of the initial bite.

Originating from the pioneering work on the auscultation of the masticatory system (system related to chewing) done by Brenman [19] and Watt [20] the stability of occlusion and has been assessed in the field of oral rehabilitation by analyzing teeth contact sounds (gnathosonic analysis) [21]. Similarly the sounds produced by the temporomandibular joints during jaw opening and closing movements have been studied regarding joint dysfunction [22]. It is not expected that these sound sources provide a audible contribution to chewing of food materials in healthy subjects. However, these studies provide information regarding sound transducer types and mounting position that may be usable also for the analysis of chewing sounds. Recent investigations [23, 21] evaluated measurement methodology, applicable transducers types and positions.

2 Methodology

This section will give an overview of our approach. It is important to note that, as described in the introduction, we consider the sound analysis to be just one part of a larger dietary monitoring system. This means that sound analysis is not meant to solve the entire dietary monitoring problem by itself. Instead the goal of our work is to demonstrate that a significant amount of useful information *that is difficult to obtain through other means* can be extracted from chewing sound analysis. Furthermore, the question how it can be expected to interact with other context information is an important research question pursued by our group (although it is not the focus of this paper).

2.1 Approach

Nutrition intake can be coarsely divided into three phases: fracturing (tearing) the food mainly with the incisors, chewing of the pieces and swallowing of the bolus. Ultimately, all three phases should be analyzed since the bolus formation process differs for characteristic food materials [24], e.g. a dry potato chip differs in structure, fluid compartments and chewing from cooked pasta. Initial bites may have more distinctive properties [18], but occur less often and are not available for all food types. A combination of fracture sound and bolus production process features may permit the acoustic detection of food products.

In this paper, we concentrate on the longest phase. Therefore we have chosen to analyze the sound of normal chewing cycles, i.e. beginning after intake of the food piece up to and excluding swallowing of the bolus. We stopped with analyzing the sound when the amplitude level decayed to approximately 5dB above the noise level.

Fig. 1 illustrates the overall structure of our approach. It consists of three main steps: signal acquisition, chewing segment identification and food type classification.

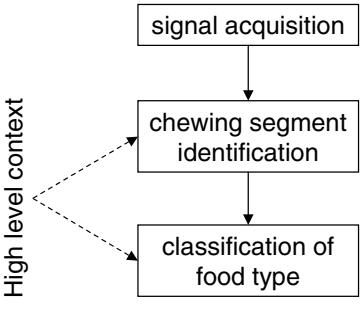

Fig. 1. Approach to the analysis of chewing sounds

The challenge of signal acquisition is to identify a microphone position that combines good amplitude levels for the chewing sounds, with good suppression of other sounds at a location that is comfortable and socially acceptable to the user.

For chewing segment identification this paper considers only sound-related means. In particular, we investigate a classifier that can distinguish between a broad range of chewing sound and various speech/conversation sounds. In a wearable computing environment, other means are possible. E.g., food intake is usually accompanied by moving the arm up and bringing the hand close to the users mouth. The lower arm is then pointing away from the earths center of gravity; something which can easily detected by an accelerometer mounted on the users wrist. However, the user can perform similar movements for other activities (e.g. scratching his chin) so other information from sensors in the

environment might be needed (e.g. location information that the user is in the kitchen or the dining room).

Once a segment is classified as being a chewing sound, the type of food needs to be identified. Again, we focus on the audio analysis of the chewing sound. In doing so, we do not aim to be able to pick any of the thousands of possible food types. This would clearly be unrealistic. Instead we assume (1) that we have a certain prior knowledge about the type of foods that are relevant to the particular situation and (2) that often it is sufficient to just be able to identify a general type of food or be able to say "could have been XY". The first assumption is not as far fetched as it might sound. The intelligent refrigerator/cardboard that knows what food is inside and what has been taken out (e.g. through RFID) is the prototypical ubiquitous application. In a restaurant credit card information or an electronic menu could be used to constrain the number of possibilities. Additionally, people have certain fairly predictable eating habits. The second point relates to the type of application that is required. As stated in the introduction, the system does not need be fully automated to be useful and to be an improvement over current 'manual' monitoring. Thus it is perfectly sufficient if at the end of the day the system can remind the user that for example "at lunch you had something wet and crisp (could have been salad) and some soft texture stuff (spaghetti or potatoes)" and asks him to fill in the details. From the above considerations we concentrate our initial work on being able to distinguish between a small set of predefined foods and on the distinction between certain food classes.

2.2 Experiments

The evaluation of all methods described in the remainder of the paper has been performed using the following experimental setup.

Test subjects: Four subjects (2 female, 2 male, mean age 29 years) were instructed to eat different food products normally, with the mouth closed during chewing. In this way the chewing phase of the nutrition cycle is covered: Beginning after intake of the food piece up to swallowing of the bolus (see Sec. 2.1).

By restricting our experiments to the chewing phase, we ensure that the recognition works solely on chewing. Specifically, we exclude swallowing and tearing sounds since these phases have different acoustic characteristics. Fracturing (tearing) and swallowing sounds are regarded as additional source of information and may be analyzed independently. Since these events are not occurring at the same high frequency than chewing, they are considered less relevant.

The subjects had no denture, no acute teeth or facial pain and no known history of occlusion or temporomandibular joint dysfunction. Furthermore none of the subjects expressed a strong dislike of any food product in this study.

Test objects: The food products shown in Table 1 have been selected since they imitate typical components in a meal or daily nutrition. The food groups reflect the acoustic behavior during chewing and not their nutrition value. They can be

simply reproduced with a high fidelity. Furthermore some of the crisp-classified products have been referenced in texture studies before: Potato chips [17] and apples [18]. Beside the dry-crisp and wet-crisp categories, a third acoustic group of "soft texture" foods have been included: Cooked pasta and cooked rice.

Table 1. Details for the food products and categorization

Food product	Food group	Product/Ingredients/Preparation
Potato chips	dry-crisp	Zweifel, potato chips (approx. 3cm in diameter)
Apple	wet-crisp	type "Jonagold" and "Gala" washed, cut in pieces, with skin
Mixed lettuce	wet-crisp	endive, sugar loaf, frisée, raddichio, chicory, arugula
Pasta	"soft texture"	spaghetti (al dente)
Rice	"soft texture"	rice without skin

Initial evaluation of the sound data showed that the rice recordings were smallest in amplitude of all recorded foods. The potato chips produced the highest amplitude for all subjects. Fig. 5 illustrates a typical waveforms recorded for apples.

Table 2 depicts the inspected sound durations for the food products from all subjects. The number of single chews is the number given by the single chew detection algorithm explained in Sec. 5.1. The single chews per chewing sequence reflects the authors' experience that usually potato chips are destruced with only a few chews, whereas pasta or lettuce require several chews to masticate properly.

Table 2. Statistics of the acquired and inspected sounds for all food products

Food product	Time recorded and inspected	No. of chewing sequences	Detected No. of single chews	Single chews per chewing sequence
Potato chips	677 sec	179	979	5.5
Apple	1226 sec	245	1538	6.3
Mixed lettuce	1054 sec	152	1691	11.1
Pasta	630 sec	74	1290	17.4
Rice[a)]	240 sec	-	-	-
Total	3827 sec	650	5498	

[a)] omitted because of small amplitude, see Sec. 4

Test procedure: A electret condenser microphone (Type Sony ECM-C115) was placed in the ear canal as described in Sec. 3. After positioning, the microphone fixation was checked to avoid interference between movements of the jaw and

the microphone in the ear canal. A second microphone of the same type was used at collar level, at the side of the instrumented ear, as reference to detect possible environmental sounds during inspection. The waveforms were recorded at a sampling frequency of 44.1 kHz, 16 bit resolution.

All products were served on a plate. Cutlery was used for the mixed lettuce, pasta and rice. Subjects were instructed to take pieces, small enough to be ingested and chewed at once, as described above. The temperature of pasta and rice was cold enough to allow normal chewing.

3 Positioning of the Microphone

Sound produced during the masticatory process can be detected by air- and bone-conduction. Frequency analysis of air-conduced sounds from chewed potato chips showed spectral energy between zero and 10 kHz [10] although the frequency range with highest amplitude for various crisp products are in the range of 1 kHz−2 kHz [25]. Bone-conducted sounds are transmitted through the mandibular bones to the inner ear. The soft tissue of mouth and jaw damp high frequencies and amplify at the resonance frequency of the mandible (160 Hz) when chewed with closed mouth [11].

Condenser or dynamic microphone transducers have been used in texture studies literature at various places with the goal to detect and reproduce human perception. Mainly the following positions were evaluated: In front of the mouth [9,10], at the outer ear above the ear canal [13], a few centimeters in front of the ear canal opening [12], pressed against the cheek [9, 12] or placed over the ear canal opening [9, 18]. Gnathosonic studies used a stereo-stethoscope technique [20] and microphones [11] at the forehead or over the zygoma [26]. More recently a method using head-phones with the microphones positioned over the ear canal opening has been proposed [21].

Several positions for the microphone have been evaluated for this study as indicated in Table 3. This list includes some of the positions used in previous

Table 3. Evaluated microphone positions

Microphone	Position
1	Inner ear, directed towards eardrum (Hearing aid position)
2	2cm in front of mouth (Headset microphone position)
3	At cheek (Headset position)
4	5cm in front of ear canal opening (Reference position for audible chewing sounds)
5	Collar (Collar microphone position)
6	Behind outer ear (Hidden by the outer ear, used by older hearing aid models)

work. The evaluation of ubiquitous positions, not hindering the user's perception was emphasized. To this end, positions 1, 5 and 6 are favorable because their implementation can be hidden in human anatomy or in cloths.

Potential artifacts introduced by daily use could interfere significantly with the microphone function. This may affect position 5 since it has the disadvantage of being hidden under cloths or disturbed by cloth sounds. Position 1 has the advantage of being less affected by loud environmental noises since it is embedded directly into the ear canal: With a directional microphone oriented towards the eardrum, the intensity of any noise from the environment is reduced.

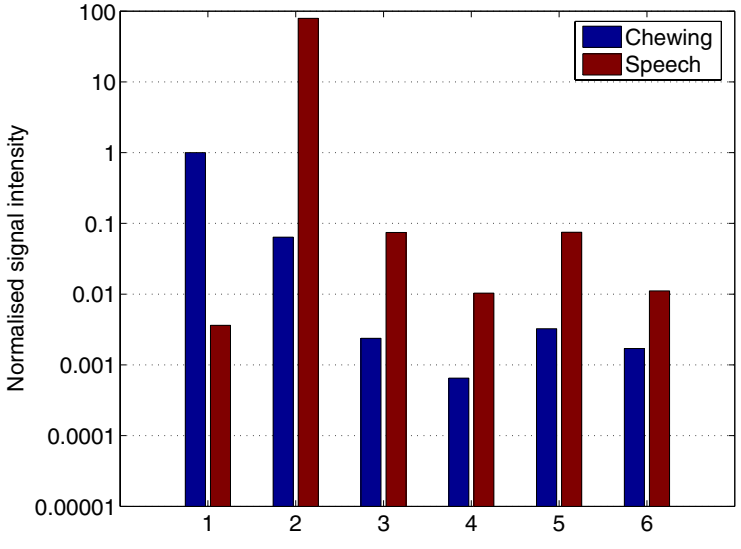

Fig. 2. Signal intensity of different microphone positions (see Table 3)

The position of the microphone was evaluated while a subject was chewing potato chips and while the subject was speaking. The mean amplitude perceived at position 1 was used as reference for normalization. Fig. 2 depicts the relation of the signal amplitude intensity shown on a logarithmic scale. It can be seen clearly that position 1 not only has the highest intensity for chewing sounds but it is also the only position with chewing sound intensity higher than speech intensity. Therefore for all further measurements position 1 was used.

A microphone at position 1 does not need to hinder the person, as modern hearing aids prove. Applicable microphones could be very small and combined with an earphone be used for other applications, e.g. mobile phones. For example, modern hearing aids already operate with a combined microphone/earphone.

4 Chewing Segment Identification

The identification of chewing segments in a continuous sound signal can be regarded as a base functionality and hence is of high importance for the detailed analysis of the masticated food type. We see mainly two different methods based on audio signal processing.

A: Intensity of Audio Signal: In an environment, like a living room, with background music playing or in a quiet restaurant, the chewing sound picked up in the inner ear is much louder than a normal conversation or background music. This is indicated in the sample recording shown in Fig. 3.

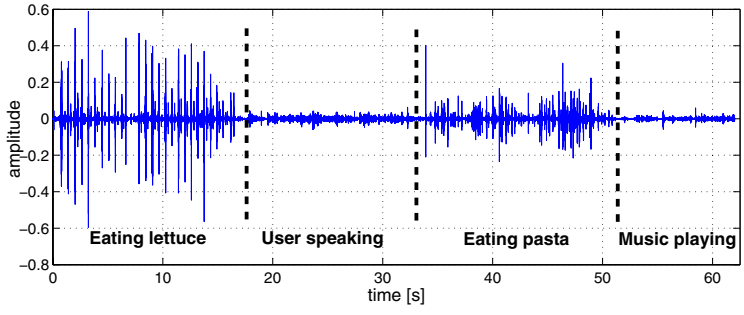

Fig. 3. Chewing sound and speech recording in a room with background music

B: Chewing Sound – Speech Classifier: Despite the general suppression of the speech signal, loud speech can at times develop amplitude peaks similar to chewing signals. Therefore it is necessary to be able to separate these two classes. This is achieved by calculating audio features from a short signal segment of length t_w, averaging the features over N_{avg} segments and then finally classifying them with a previously trained classifier [27].

Features: We used features that are popular in the area of speech, audio and auditory scene recognition [28, 29, 30]. In the temporal domain, those were zero-crossing rate and fluctuation of amplitude. Frequency domain features were evaluated based on a 512-point Fast Fourier Transformation (FFT) using a Hanning window. Here, the features included: frequency centroid, spectral roll-off point with the threshold of 0.93, fluctuation of spectrum and band energy ratio in 4 logarithmically divided sub-bands. Additionally 6 cepstral coefficients (CEP) were evaluated. Both time and frequency domain features were evaluated on a window of $t_w = 11.6$ ms. No overlap between the windows was used.

The features were averaged over N_{avg} windows to improve the recognition results. This method helped to bridge pause gaps between the chewing sounds. These gaps vary between 100 ms and 600 ms depending on the chewed material

and the progression of decomposition (see Fig. 5). Longer pauses may be observed at the beginning of a chewing sequence for larger food pieces as well as before and after partial bolus swallowing.

Classifiers: A C4.5 decision tree classifier from the Weka Toolkit [31] was trained with the aforementioned features. The classifier was 10-fold cross-validated on a two class data set. The first class contained all food products as specified in Table 2 except cooked rice. Rice was excluded since individual classification of food products against speech signals showed weak results for rice. This was expected from the low signal-noise ratio of the rice sounds. The second class included various speech signal segments from several speakers as well as conversation of test subjects and the authors.

Since the accuracy of a classifier depends on the class distribution, the ROC curve (Receiver Operating Characteristic) is presented instead (see Fig. 4). ROC curves help to visualize classifier performance over the whole range of frequency of occurrence [32]; the best classifier is the one to the top-left corner. This is useful in our case since the number of occurrences of speech and chewing sounds may vary and may not be known beforehand. Clearly, the classifier that uses the CEP features dominates. This was expected since the CEP features help to pick out speech sequences. Furthermore, the number N_{avg} of averaging frames was varied. We found that the highest recognition rates can be achieved if N_{avg} is chosen so that the features are at least averaged over one single chew which takes about one second. In our case this occurs if $N_{avg} > 1\ \mathrm{sec}/t_w = 86.2$.

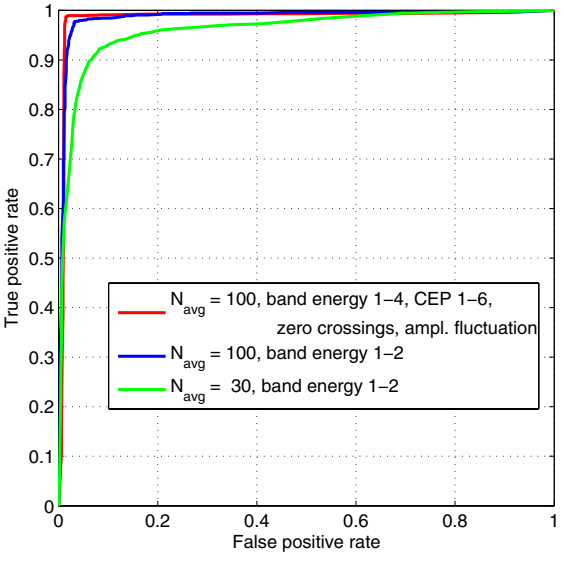

Fig. 4. ROC curve for chewing sounds (positives) and speech sounds (negatives)

5 Discrimination of Foods Products

5.1 Isolation of Single Chews

First trials in separating different food products with the same methods as in the previous section (i.e. calculating features over a large window) produced recognition rates around 60%. The reason for this is mainly due to the rather long pause between single chews, which produces the same audio signature for all food items.

To overcome this problem we have looked in more detail at the temporal structure of a typical chewing sequence (see Fig. 5). It can be seen that the audio signal of one chew is mainly composed of four phases: The closing of the mandible to crush the material, a small pause, the opening of the mandible in which material that stick to the upper and lower teeth is uncompressed, and again a pause. The timing between those phases is given mainly by the mechanical properties of the food and the physical limitations of the mandible. All test subject showed almost the same timing for the same food, with the exception of a longer or shorter pause in phase 4 (fast/slow eater). The four phases are very well distinguishable in crispy food, in softer food like pasta the phases tend to merge. Still, the pause in phase 4 and the increase in amplitude at the beginning of phase 1 remain.

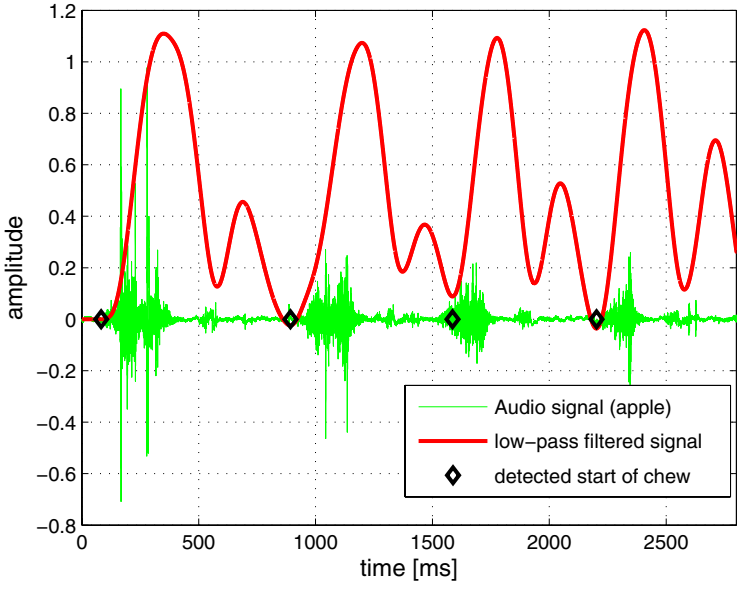

Fig. 5. Sample sound signal observed for chewing an apple

A relatively simple algorithms helps us the detect the beginning of each chew. The short-time signal energy in a 20 ms window is compared to a energy

threshold and the resulting signal is set to 1 if the short-time signal energy is larger than the threshold and to 0 otherwise. The resulting signal is low-pass filtered with a 4th order butterworth filter. We found that a filter with a 3dB cut-off frequency of 4 to 5 Hz reliably responds to the pause in phase 4 while filtering out the shorter pause in phase 2. With help of the hill climbing algorithm the beginning of each chew is detected as shown in Fig. 5. We found that this algorithm can detect the start point of about 90% of all chews while producing only very little insertions.

5.2 Classification

Once the audio signal is segmented into single chews, the segments are classified using the same procedure as in Sec. 4. Several features were applied to a short window that was consecutively shifted. We found that a 11.6 ms window with a shift of 8.7 ms works best for our sound classes. The most promising features were: zero crossing rate, band energy ratios, fluctuation of amplitude, fluctuation of spectrum and bandwidth. The features were further averaged over the length of a single chew. The length of a single chew was used as an additional feature and helped to improve the recognition rate of especially the pasta, since soft-texture foods have shorter durations of chews. The features were then 10-fold cross-validated with a C4.5 decision tree classifier. Recognition rates range around 66% to 86% and the corresponding confusion matrix is listed in Table 4.

Table 4. Confusion matrix for single chews

a	b	c	d	← classified as	Accuracy
669	170	25	115	a = Chips	68.34%
183	1024	41	290	b = Apple	66.58%
25	39	1112	114	c = Pasta	86.20%
125	293	95	1178	d = Lettuce	69.66%

Since the material inside the mouth can not change between single chews, a majority decision over a whole chewing cycle was performed. This measure resulted in an increase of recognition rate of 15 to 20% as shown in Table 5. It can be seen that there is some confusion between apple and lettuce which can

Table 5. Confusion matrix for chewing cycles

a	b	c	d	← classified as	Accuracy
156	12	1	10	a = Chips	87.15%
24	198	1	22	b = Apple	80.82%
0	0	74	0	c = Pasta	100.00%
4	21	0	127	d = Lettuce	83.55%

be explained by them belonging into the same food category (see Table 1) and therefore having similar mechanical properties.

6 Conclusion and Future Work

6.1 Conclusion

The work presented in this paper has proven that chewing sound analysis is a valuable component for automated dietary monitoring systems. Specifically we have shown that:

1. A microphone location inside the ear can acquire good quality chewing sounds while suppressing many other sounds originating inside the oral cavity such as speech. At the same time it is a location that has been proven to be acceptable to users in other applications (e.g. hearing aids, headsets). Applicable microphones could be very small, not hindering the normal perception. Moreover, a combination of microphone and earphone for shared use with other applications, e.g. a mobile phone, could be employed.
2. Chewing sounds can be reliably separated from the main sound source inside the mouth cavity: speech.
3. Individual chews can be isolated and partitioned into phases with a simple low pass filter based algorithm
4. Audio analysis can be used to distinguish between a small predefined set of different food types as for example found in a single meal.

The food groups introduced in the experiments reflect the acoustic behavior during chewing and not their nutrition value. The results show, that our approach is not limited to a specific group of foods. Moreover, it is possible to discriminate foods from the same group. The actual nutrition value can be derived either precisely from other monitoring components, e.g. RFID tags of packages, or as an estimate from a generic food database.

An important aspect of our work is the fact that information about the specific type of food which is being chewed is very difficult to derive using other sensor modalities. The only alternative we could think of is video analysis of the items inserted into the mouth. While theoretically feasible it has many problems of its own, in particular sensitivity to light conditions and background clutter as well as large computational complexity.

Overall the results presented in this paper provide crucial groundwork for further development that, we believe, will lead to complete automated dietary monitoring systems. Within the scope of the EU funded MyHeart project we aim to have first versions of such a system within the next two to three years. Additionally, points 1 and 2 have implications beyond dietary monitoring as they allow a fairly accurate recognition of the fact that the user is eating. This in itself is an important context information.

6.2 Future Work

On the sound analysis the next steps that we will undertake are:

1. Modeling temporal evolution of the signal from individual chews with hidden Markov models to further increase the recognition rates and allow similar food types to be distinguished.
2. Modeling the temporal evolution of the individual chewing signals over an entire chewing cycle to extract food type specific parameters. This shall include the number of individual chews needed, their length and the evolution of the sound intensity.
3. Performing studies about the robustness of the system by adding controlled levels of noise.
4. Performing more studies with more, different food types.
5. Performing studies to determine how the recognition performance degrades with increasing number of food types that need to be differentiated.
6. Using a hierarchical approach with an initial classification of the category (dry crisp, wet crisp etc.) and then a category specific algorithm for further recognition, to overcome the above limitation.

Furthermore, other components of a dietary monitoring system will also be investigated. In particular, we will look at the detection of swallowing motion with collar electrodes, analyze the hand motions related to food intake and integrate high level context information relevant to eating habits into the system.

References

[1] Mynatt, E., Melenhorst, A.S., Fisk, A.D., Rogers, W.: Aware technologies for aging in place: understanding user needs and attitudes. In: IEEE Pervasive Computing. Volume 3. (2004) 36–41
[2] Gellersen, H.W., Schmidt, A., Beigl, M.: Multi-sensor context-awareness in mobile devices and smart artifacts. Mobile Networks and Applications **7** (2002) 341–351
[3] Philipose, M., Fishkin, K., Perkowitz, M., Patterson, D., Fox, D., Kautz, H., Hahnel, D.: Inferring activities from interactions with objects. IEEE Pervasive Computing **3** (2004) 50–57
[4] Mihailidis, A., Carmichael, B., Boger, J.: The use of computer vision in an intelligent environment to support aging-in-place, safety, and independence in the home. IEEE Transactions on Information Technology in Biomedicine **8** (2004) 238–247
[5] Römer, K., Schoch, T., Mattern, F., Dübendorfer, T.: Smart identification frameworks for ubiquitous computing applications. In: PerCom 2003: Proc. of the First IEEE Int'l Conference Pervasive Computing and Communications. (2003)
[6] Beigl, M., Gellersen, H.W., Schmidt, A.: MediaCups: Experience with design and use of computer-augmented everyday artefacts. Computer Networks, Special Issue on Pervasive Computing **35** (2001) 401–409
[7] Vickers, Z., Christensen, C.: Relationships between sensory crispness and other sensory and instrumental parameters. Journal of Texture Studies (1980) 291–307
[8] Vickers, Z.: Relationships of chewing sounds to judgements of crispness crunchiness and hardness. Journal of Texture Studies (1981) 121–124

[9] Drake, B.: Food crushing sounds. an introductory study. Journal of Food Science (1963) 233–241
[10] Lee, W., Schweitzer, G., Morgan, G., Shepherd, D.: Analysis of food crushing sounds during mastication: Total sound level studies. Journal of Texture Studies (1990) 156–178
[11] Kapur, K.: Frequency spectrographic analysis of bone conducted chewing sounds in persons with natural and artifcial dentitions. Journal of Texture Studies (1971) 50–61
[12] Dacremont, C., Colas, B., Sauvageot, F.: Contribution of air- and bone-conduction to the creation of sounds perceived during sensory evaluation of foods. Journal of Texture Studies (1991) 443–456
[13] Vickers, Z.: Sensory acoustical and force? Deformation measurements of potato chip crispness. Journal of Texture Studies (1987) 138–140
[14] Szczesniak, A.: Texture: Is it still an overlooked food attribute? Food Technology (1990) 86–95
[15] AlChakra, W., Allaf, K., Jemai, A.: Characterization of brittle food products: Application of the acoustical emission method. Journal of Texture Studies (1996) 327–348
[16] Edmister, J., Vickers, Z.: Instrumental acoustical measures of crispness in foods. Journal of Texture Studies (1985) 153–167
[17] Vincent, J.F.V.: The quantification of crispness. Journal of Science in Food Argiculture (1998) 162–168
[18] DeBelie, N., De Smedt, V., J., D.B.: Principal component analysis of chewing sounds to detect differences in apple crispness. Journal of Postharvest Biology and Technology (2000) 109–119
[19] Brenman, H., Weiss, R., Black, M.: Sound as a diagnostic aid in the detection of occlusion discrepancies. Penn Dental Journal (1966)
[20] Watt, D.: Gnathosonics - a study of sound produced by the masticatory mechanism. Journal of Prosthetic Dentistry (1966)
[21] Prinz, J.: Computer aided gnathosonic analysis: distinguishing between single and multiple tooth impact sounds. Journal of Oral Rehabilitation (2000) 682–689
[22] Widmalm, S., Williams, W., Zengh, C.: Time frequency distribution of tmj sounds. Journal of Oral Rehabilitation (1991)
[23] Tyson, K.: Monitoring the state of occlusion - gnathosonics can be reliable. Journal of Oral Rehabilitation (1998) 395–402
[24] Hutchings, J., Lillford, D.: The perception of food texture - the philosophy of the breakdown path. Journal of Texture Studies (1988) 103–115
[25] Brochetti, D., Penfield, M., Burchfield, S.: Speech analysis techniques: A potential model for the study of mastication sounds. Journal of Texture Studies (1992) 111–138
[26] Watt, D.: Gnathosonics and Occlusal Dynamics. Praeger New York (1981)
[27] Stäger, M., Lukowicz, P., Perera, N., von Büren, T., Tröster, G., Starner, T.: SoundButton: Design of a Low Power Wearable Audio Classification System. In: ISWC 2003: Proc. of the 7th IEEE Int'l Symposium on Wearable Computers. (2003) 12–17
[28] Li, S.Z.: Content-based audio classification and retrieval using the nearest feature line method. IEEE Transactions on Speech and Audio Processing **8** (2000) 619–625
[29] Li, D., Sethi, I., Dimitrova, N., McGee, T.: Classification of general audio data for content-based retrieval. Pattern Recognition Letters **22** (2001) 533–544

[30] Peltonen, V., Tuomi, J., Klapuri, A., Huopaniemi, J., Sorsa, T.: Computational auditory scene recognition. In: IEEE Int'l Conf. on Acoustics, Speech, and Signal Processing. Volume 2. (2002) 1941–1944
[31] Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann (1999)
[32] Provost, F., Fawcett, T.: Analysis and visualization of classifier performance: Comparison under imprecise class and cost distributions. In: KDD '97: Proc. of the 3rd Int'l Conference on Knowledge Discovery and Data Mining. (1997) 43–48