# Evaluation of Algorithms for Chew Event Detection

Sebastian Päßler*, Wolf-Joachim Fischer
Fraunhofer IPMS
Maria-Reiche-Straße 2
01109 Dresden
*Tel.: +49(0) 351/8823-212

*corresponding author: Sebastian.Paessler@ipms.fraunhofer.de

## ABSTRACT

Analyzing food intake behavior is necessary to prevent obesity and overweight. Detecting and counting chewing strokes is an elementary part of this analysis. In our project, sounds of food intake were recorded using a microphone in the outer ear canal. The records contained sounds of 51 healthy subjects chewing 8 types of food. We evaluated seven different algorithms to detect chew events in sound records. Results of the automated detection were compared to manual annotations. Best performances (precision and recall over 76 %) were achieved by detecting chew events in six different frequency bands and fusing these results. With this method for counting the number of chews, an important step towards the estimation of bite weight has been done.

## Categories and Subject Descriptors

J.3 [**Life and Medical Sciences**]: Health – *food intake monitoring, chewing sound, on-body signal processing.*

## General Terms

Algorithms, Measurement, Performance.

## Keywords

Food intake sound, chew event detection, mobile healthcare.

## 1. INTRODUCTION

There is a growing need for individuals to plan and evaluate their food intake in order to maintain a healthy diet. Unbalanced diet and excessive eating lead to various kinds of health risks such as cardiovascular disease, diabetes mellitus and cancer. A significant increase of these kinds of nutrition related diseases during the past decades was reported by the World Health Organization (WHO) in 2006 [12]. Approximately 1.6 billion adult people worldwide have been stated to be overweight. At least 400 million of them were classified to have reached an obese state. This challenge has high impact especially in the population of elderly individuals under the conditions of ambient assisted living. Nursing auxiliaries and physicians regularly check the health state of these people and protocol their food intake by interviewing them once a day.

The accuracy of this procedure depends on the skills of the interviewers and the mental conditions of the patient. Estimation errors of the amount of consumed food of up to 50 % are not unusual [10]. Additionally, snacking between the meals is often forgotten to be reported.

A wearable device for precise and timesaving logging of human food intake data may be of great benefit in the task of analyzing eating behavior. Analyzing and classifying the sounds generated during the human food intake process seems to be a promising way to protocol food intake. Several research groups developed systems to get information of the food intake process by sound analysis [3, 6, 7, 11]. Body worn devices share the benefit of being independent from the eating location and lighting conditions.

In [2], Amft et al. investigated the relation of the bite weight of consumed food and several parameters of the food intake process, e.g. the number of chewing strokes. They concluded that bite weight prediction using this kind of information results in estimation errors in the range of self-reporting. Thereby, the automated prediction procedure can drastically reduce the user's effort on data logging.

The performance of a bite weight prediction algorithm strongly depends on the accuracy of the spotting of chew events in the sound data. Several authors have faced the task of spotting chew events [2, 6, 11]. In this paper, we want to evaluate the performance of different algorithms for this task on a database of records. Additionally, we investigate the usability of approaches from the field of rhythm detection in music for chew event spotting.

This paper is organized as follows: In section 2, we present related studies on the task chew event detection. An Algorithm from the field of music analysis is presented here as well. Details of the wearable sensor system for food intake monitoring developed by our group are given in section 3. Section 4 explains the chew event spotting algorithms under investigation. Section 5 presents the methods and results of the evaluation experiments. Discussion and conclusion are drawn in section 6 and 7, respectively.

## 2. RELATED WORK
### 2.1 Studies on food intake sound monitoring

In 1963, Drake investigated acoustics of chewing hard and crisp food in a pioneer study [5]. He demonstrated that chewing of each food type generates characteristic sounds. There are several approaches to analyze chewing sounds using wearable devices in order to protocol food intake behavior in the past decades: Amft developed a wearable earpad sensor with a microphone inside of a headphone housing [3]. Using a device like this, Amft et al. derived the boundaries of chew events from the raw sound data and classified the type of consumed food. Sound data were recorded

from eight participants chewing three types of crispy food. Chew event spotting and food classification was done using feature similarity search. Shuzo et al. [11] recorded chewing sounds using a wearable recorder with miniature bone conduction microphones. The microphones were placed inside of the outer ear canals. The authors spotted periods of food intake, classified consumed food according to its texture and counted mastications (i.e. chewing strokes). Detection and counting of chewing strokes was based on filtering the absolute values of the signal with a low-pass filter. A cut-off frequency in the range of 0.5 to 5 Hz was used. Chewing strokes were spotted by detection of maximum values in the filtered signal. The recording device of Nishimura and Kuroda [6] consisted of a microphone integrated into a Bluetooth headset. Using it, periods of food intake could be detected by spotting "chew-like" signals and a classification of "crisp" and "non-crisp" food types was carried out. "Chew-like" signals were detected as rhythmic sounds with amplitudes modulated with a frequency of not more than 2 Hz. For this, the signal was divided into subsequent frames and the log energy was calculated in each frame. The resulting log energy signal was low-pass filtered with a cut-off frequency of 2 Hz. Regression coefficients of 9 adjacent frames on the filtered signal and zero crossings of these coefficients were calculated. Finally, chewing strokes were marked when the filtered signal was larger than a threshold and regression coefficients had negative slope when crossing zero. The "chew-like" signals were verified using models based on Mel-frequency cepstral coefficient (MFCC) features. A wearable sensor system for food intake monitoring has been designed by our group [7]. The system is designed especially for elderly people and explained in more detail in section 3. We recorded food intake sounds from users at a sampling rate of 11,025 Hz. Using these records, we were able to classify different types of consumed food based on their chewing sounds [7].

## 2.2 Studies on beat detection in music

A common task in music analysis is the so called beat detection task. Beats generated by drums or bass guitars in music recordings are detected by special designed algorithms. Rhythm analysis can be based on such a beat detection algorithm. One example of beat detection algorithms is presented here. It was based on the "Beat This" project published online in [4]. Six spectral bands were extracted from the spectrum of the Fast Fourier Transform (FFT) of a whole record. Time signals were calculated by inverse Fast Fourier Transform (IFFT) of each spectral band. Absolute values of these signals were calculated. Subsequently, they are low-pass filtered by convolution with the function of a half-Hanning window. The resulting signal is differentiated and rectified. The dominating peaks of these rectified signals demark the occurrence of beats in music signals.

## 3. WEARABLE SENSOR SYSTEM

In order to analyze food intake behavior in elderly people, we designed a sensor system with two tiny electret microphones (FG-23329-CO5, Knowles Acoustics) integrated in a hearing aid case [7]. Figure 1 shows the application part of the system. One microphone ("in-ear microphone") was placed in the outer ear canal of the user to record chewing sounds. It was protected from ear wax by a silicon tip. The decision for this application place was made based on investigations of Amft et al. [1], who found the outer ear canal best to record chewing sounds from the acoustical point of view. The second microphone ("reference microphone") was placed in the hearing aid case above the outer ear. It recorded
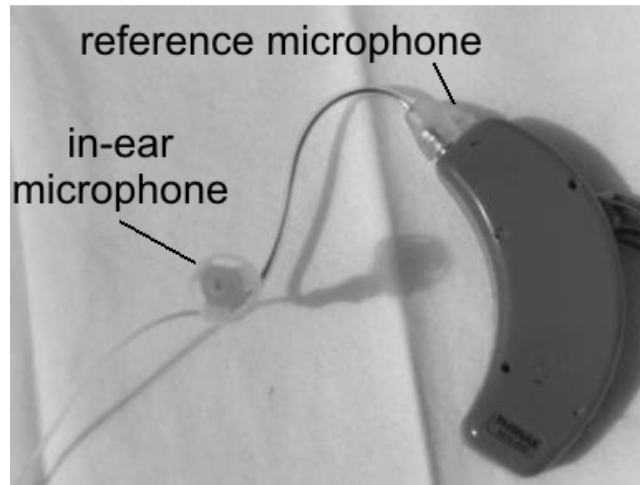


**Figure 1. Hearing aid package with both integrated microphones.**

environmental sounds not generated by the body. The signals of both microphones were amplified and filtered in an analogous stage. Both signals were sampled synchronously by a standard sound card of a notebook computer. A sampling rate of 11,025 Hz was used at a quantization of 16 bit.

## 4. CHEW EVENT DETECTION ALGORITHMS

We investigated seven different algorithms for spotting of chew events. These algorithms are presented in this section.

## 4.1 Maximum sound energy algorithm

Chew events were spotted by this algorithm when signal energy reached a maximum value, thereby extending a threshold. For this, the sound signal was divided into subsequent frames of 23
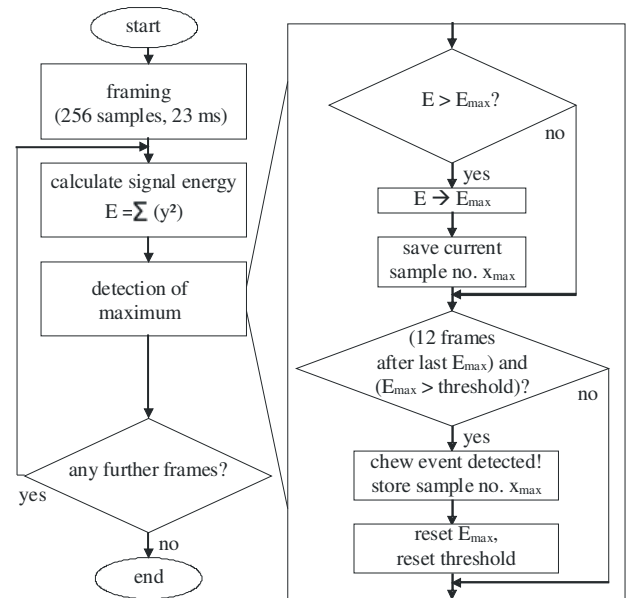


**Figure 2. Workflow of the maximum sound energy algorithm. Symbolism: E is the signal energy, y is the current signal sample value, $E_{max}$ is the latest detected signal energy maximum, $x_{max}$ is the sample number of $E_{max}$.**

ms (i.e. 256 samples) without overlap. Signal energy was calculated as the sum of the squared signal values of each sample in a frame.

Maxima of the signal energy in the sequence of the frames of a record were detected as follows: A maximum was spotted if signal energy of each of the subsequent 12 frames was lower than the value of the current frame. Only if the signal energy maximum exceeded a threshold value, its occurrence was marked as chew event. Otherwise, the next signal energy maximum was evaluated. After the occurrence of a chew event, the threshold value was set to the signal energy of the 12th frame subsequent to the frame containing the chew event. These variation of the threshold value was restricted by a given minimum value.

The workflow of the maximum sound energy algorithm for chew event detection is shown in Figure 2. This algorithm is the simplest one for chew event detection under evaluation in this paper.

## 4.2 Maximum spectral band energy algorithm

Using this algorithm, the sound signal was divided into subsequent frames of 5.8 ms (i.e. 64 samples) without overlap. The signal in each frame is transformed using the FFT. The sum of the absolute spectral values was calculated for each of six spectral bands in a frame. This sum is an energy equivalent of the signal. These spectral bands were 0 to 200 Hz, 200 to 400 Hz, 400 to 800 Hz, 800 to 1600 Hz, 1600 to 3200 Hz and 3200 to 5512 Hz. For every band, maxima of the energy equivalent were calculated for the sequence of the frames in a record. This maximum detection was carried out as shown in section 4.1. Thereby, an own threshold value for the maximum energy equivalent of each spectral band was introduced. Chew events were spotted only if more than
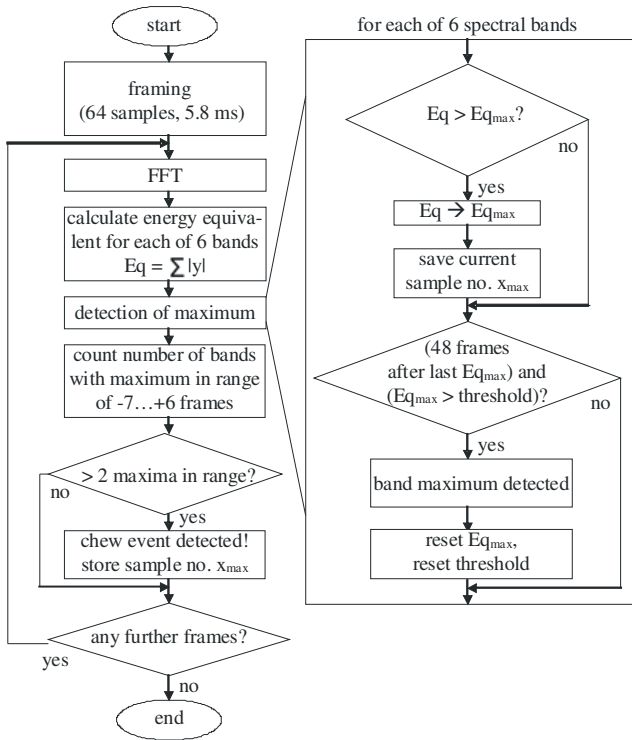
**Figure 3. Workflow of the maximum spectral band energy algorithm. Symbolism: Eq is the signal energy equivalent, y is the current signal sample value, $Eq_{max}$ is the latest detected maximum of Eq, $x_{max}$ is the sample number of $Eq_{max}$.**
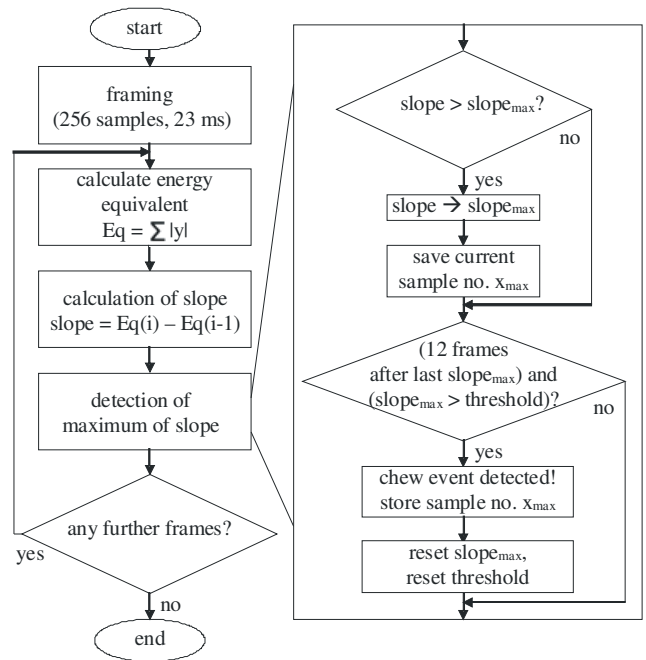
**Figure 4. Workflow of the maximum energy slope algorithm. Symbolism: Eq is the signal energy equivalent, y is the current signal sample value, i is the number of the current frame, $x_{max}$ is the sample number of the maximum slope ($slope_{max}$).**

two maxima of different spectral bands were detected in a range of 14 adjacent frames (i.e. 81.2 ms).

This algorithm takes into account the spectral properties of chewing sounds. As shown in [8], the spectra of sounds of chewing strokes are rather flat and consist of signal energy of frequencies up to 4 kHz. Sound signals with a single dominant frequency component, such as vowels in speech, are not likely to be detected by this algorithm. The workflow of the maximum spectral band energy algorithm is shown in Figure 3.

## 4.3 Maximum energy slope algorithm

Chew events are characterized by a sudden onset of chewing sounds. To detect this onset, the maximal slope values of a sound energy equivalent were spotted by the maximum energy slope algorithm. The sound signal was divided into subsequent frames of 23 ms (i.e. 256 samples) without overlap. For each frame, a sound energy equivalent was calculated as the sum of the absolute signal values of each sample. The slope of these energy equivalent values was calculated by subtracting the energy equivalent of the previous frame from that of the current one. The maxima of the slope were detected analogue to the maximum detection in section 4.1.

The maximum energy slope algorithm takes into account the sudden onset of chewing sounds. The algorithm is independent of quasi-static background noise and slowly changing sounds. However, it is expected to fail in the case of transient noise signals or as speech. The workflow of the maximum energy slope algorithm is shown in Figure 4.
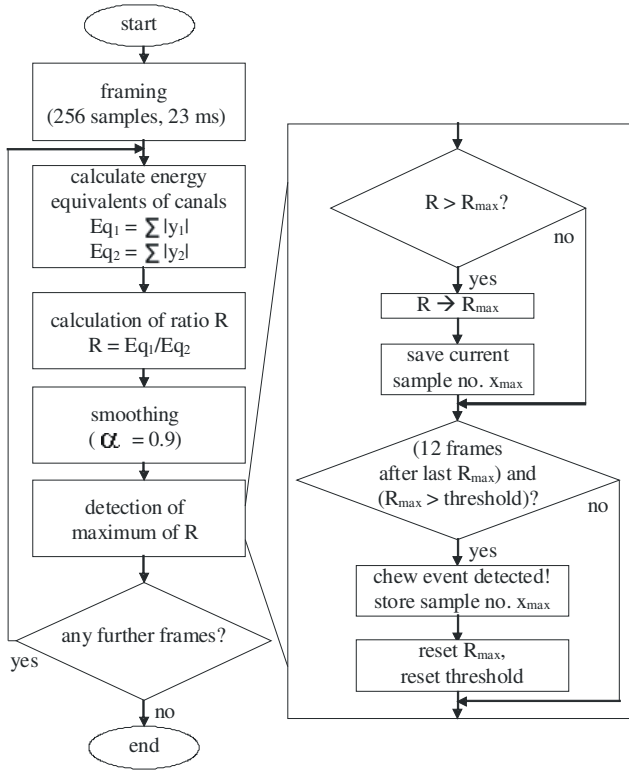
**Figure 5. Workflow of the maximum energy ratio algorithm. Symbolism: Eq1 and Eq2 are the signal energy equivalents of in-ear and reference signal, respectively, y1 and y2 are the current signal sample values of these signals, $R_{max}$ is the latest detected maximum of R, $x_{max}$ is the sample number of $Eq_{max}$.**

## 4.4 Maximum energy ratio algorithm

In our previous work [7], we showed that the ratio R of the energy equivalents of in-ear signal and reference signal could be used to distinguish signals from inside and outside of the body. This energy ratio was used in the maximum energy ratio algorithm to detect maxima in the chewing sound signal and reject sound maxima generated by the environment. The sound signal was divided into subsequent frames of 23 ms (i.e. 256 samples) without overlap. For each frame of in-ear and reference signal, respectively, sound energy equivalents were calculated as the sum of the absolute signal values of the samples. The ratio of the energy equivalents was computed by dividing the energy equivalent of the in-ear signal by that of the reference signal. To avoid division by zero, we set a limit to the energy equivalent of the reference signal by a fixed minimum value. The calculated ratio was smoothed by an adaptation factor α of 0.9 according to (1):

$$\overline{R}(i) = \alpha \cdot \overline{R}(i) + (1 - \alpha) \cdot R(i),\qquad(1)$$

where $\overline{R}$ is the smoothed ratio and i is the index of the current frame. Maxima of the smoothed ratio were detected analogue to the maximum detection in section 4.1. The workflow of the maximum energy ratio algorithm is shown in Figure 5.
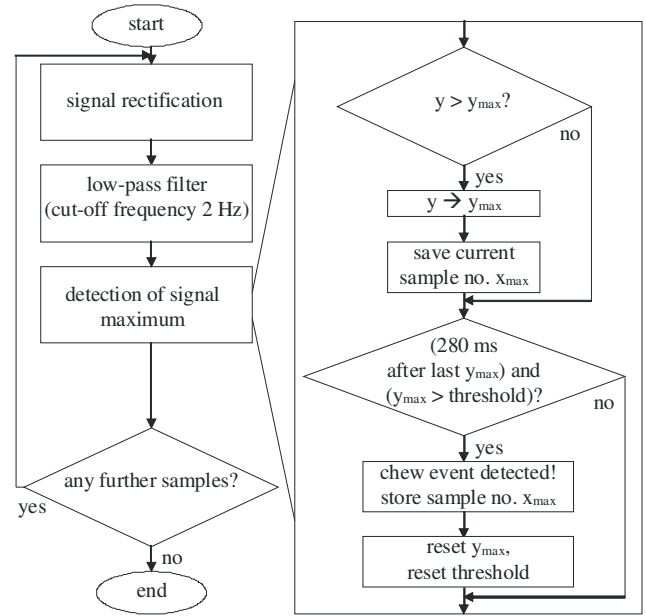


**Figure 6. Workflow of the low-pass filtered signal algorithm. Symbolism: y is the current signal sample value, $y_{max}$ is the latest detected maximum of y, $x_{max}$ is the sample number of $y_{max}$.**

## 4.5 Low-pass filtered signal algorithm

This algorithm was designed according to the algorithm for counting the number of mastications of Shuzo et al. [11]. First, the absolute value of each signal value was computed. The resulting signal is low-pass filtered with a cut-off frequency of 2 Hz. We used a Butterworth filter of fifth order. We decided to use a static cut-off frequency for ease of use. Maxima were detected in the filtered signal analogue to the maximum detection in section 4.1. In the low-pass filtered signal algorithm, maximum detection was carried out for every sample value. A maximum was spotted if the signal value of each of the subsequent 3072 samples (i.e. 280 ms) was lower than the signal value of the current sample.

Using this algorithm, a proper detection of chew events occurring with a mastication frequency less than 120/min could be expected. However, chewing with higher mastication frequencies could be rejected. Rhythmic sounds from the environment are expected to be falsely detected as chew events. The workflow of the low-pass filtered signal algorithm is shown in Figure 6.

## 4.6 "Chew-like" signal detection algorithm

The "chew-like" signal detection algorithm was developed by Nishimura and Kuroda [6]. We used it to compare the performance of our algorithms with it. In order to use it, the sound signal was divided into subsequent frames of 23 ms (i.e. 256 samples) without overlap. For each frame the logarithmic energy equivalent was computed by calculating the common logarithm of the sum of the absolute signal values. The result was filtered with a Butterworth low-pass filter of the 4th order and a cut-off frequency of 4 Hz. Mean and standard deviation of the signal values from the first nine frames of the filtered signal were computed. The threshold value for signal energy was set to seven times the sum of this mean and standard deviation value. Regression coefficients of nine adjacent frames on the filtered signal were calculated. Frames, where the regression coefficients crossed zero with a
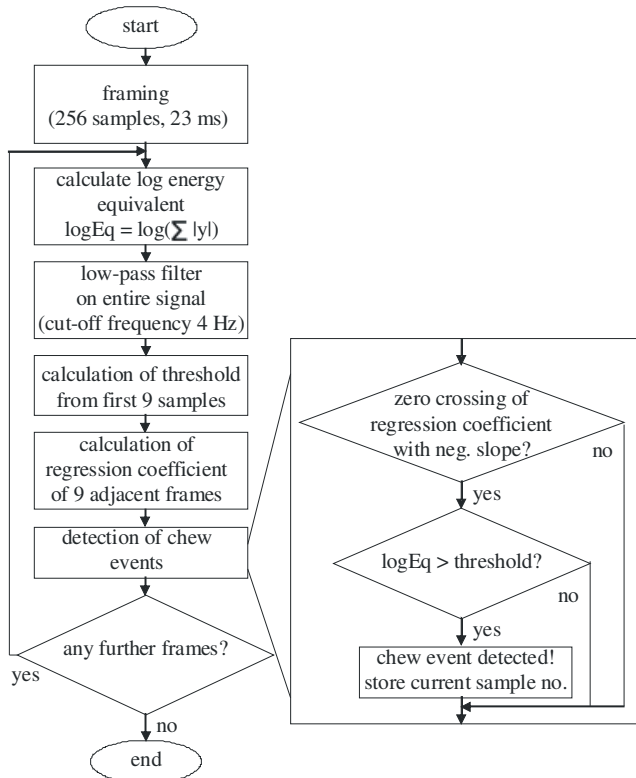
**Figure 7. Workflow of the "chew-like" signal detection algorithm. Symbolism: logEq is the logarithmic signal energy equivalent, y is the current signal sample value.**

negative slope, were marked. Finally, the markings were treated as occurrences of chewing strokes if the filtered signal was larger than the threshold value.

The workflow of the "chew-like" signal detection algorithm is shown in Figure 7. We did not use the chewing sound verification step introduced by Nishimura and Kuroda.

## 4.7 "BeatThis" algorithm

The "BeatThis" algorithm was developed based on a publication of Scheirer [9] and source code of this algorithm from [4]. The intentions of these publications were to detect beats in music recordings. The source code of [4] was based on [9]. A frequency spectrum is calculated from the whole sound record using the FFT. The spectrum was divided into sub-spectra of six spectral bands. These spectral bands were 0 to 200 Hz, 200 to 400 Hz, 400 to 800 Hz, 800 to 1600 Hz, 1600 to 3200 Hz and 3200 to 5512 Hz. Spectra of the single bands were re-transformed using the IFFT. Absolute values of the resulting signals were computed. Convolution with the function of a half-Hanning window (with a length of 500 ms) was used to smooth these signals. Finally, the first temporal slopes were calculated by subtracting each two adjacent signal values and these slopes were rectified. In [4], a calculation of the basic rhythm of a piece of music was carried out. We modified the algorithm by inserting a maximum search and dropping the rhythm calculation. Maxima of the slopes were detected for every spectral band analogue to the maximum detection in section 4.1. Chew events were spotted only if more than four maxima from different spectral bands were detected in a range of 896 samples (i.e. 81.2 ms).

The workflow of the "BeatThis" algorithm is shown in Figure 8. The "BeatThis" algorithm was designed to spot short transient
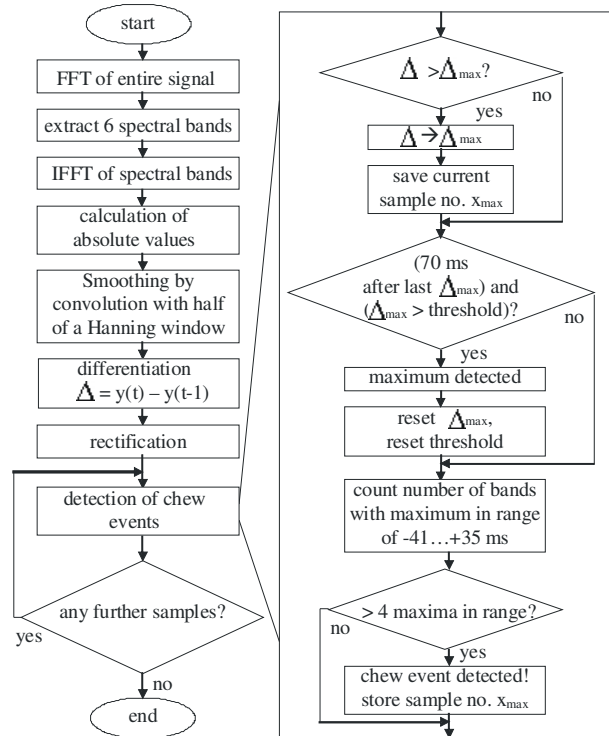


**Figure 8. Workflow of the "BeatThis" algorithm. Symbolism: y is the current signal sample value, t is the sample number of y, $\Delta$ is the time difference of y, $\Delta$ is the latest maximum of $\Delta$, $x_{max}$ is the sample number of $\Delta_{max}$.**

sounds ("beats") in music. We wanted to test whether chewing sounds could be detected as well.

## 5. EVALUATION EXPERIMENTS
## 5.1 Material and methods

Using the sensor system introduced in section 3, we recorded food intake sounds from 51 participants chewing 6 types of food. Participants were healthy subjects aged 15 to 77 years (mean: 34.8) with natural dentition. They ate 10 pieces of six types of solid food (potato chips, peanut, walnut, carrot, apple, chocolate). Food types were excluded only if participants expressed a strong dislike or were allergic of them. All participants were informed that recording can be interrupted if they feel sick or do not want to eat more pieces of food. Sound data were recorded from participants sitting at a table in a quiet office or recording room. Single pieces of food were eaten, each recorded as a single intake cycle in a separate file. Food should be chewed and swallowed as done usually. Participants were told not to talk during the record and to avoid any kind of disturbing sound. The raw sound data were labeled by trained evaluators. We achieved one label file for each sound file from this annotation procedure. Label files contained the information about onset and end of each chew or swallowing event in a sound file.

Evaluation of chew event spotting algorithms was done by comparing automatically detected chew events to the corresponding label files. If a detected chew event fell into the interval of 276 ms

around the centre of a chew event label, it was counted as true positive (TP). Any further detection in this range was counted as false positive (FP), i.e. an insertion. Detections not in this range around the centre of a chew event label were treated as false positive as well. If there was no chew event detected in this range around the centre of a chew event label, an deletion (false negative, FN) was counted. Accuracy, precision and recall values were computed as follows:

$$Accuracy = \frac{TP - FP}{TP + FN}, \qquad (2)$$

$$Precision = \frac{TP}{TP + FP}, \qquad (3)$$

$$Recall = \frac{TP}{TP + FN}. \qquad (4)$$

## 5.2 Results

Table 1 shows the performance values achieved from each algorithm on all records of a specific type of food. Overall performance values are printed in Figure 9. Best performances in terms of precision and recall could be achieved by the maximum spectral band energy algorithm and the "BeatThis" algorithm. Best prediction values were received by using the maximum sound energy algorithm and the "chew-like" signal detection algorithm. However, there are too many chew events not detected by these algorithms. Best performances were achieved for the food types potato chip, carrot and apple. Applying the algorithms on records of peanuts, walnuts and chocolate resulted in lower precision and recall values compared to the other food types. This is due to a higher number of deletion of chew events in these records.

## 6. DISCUSSION

In their publication [2], Amft et al. achieved remarkable performances of 80 % recall at 60 % - 70 % precision. These results are computed on records of eight participants consuming three different types of crispy food (potato chips, mixed lettuce and apple). We compared these results to the recognition performances of our
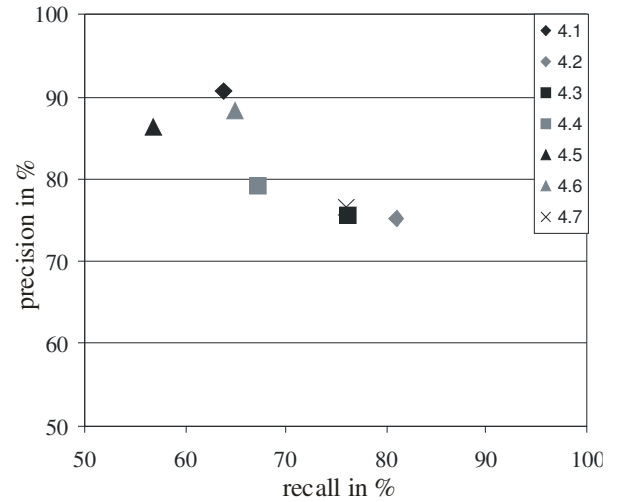


**Figure 9. Overall performance values of the evaluated algorithms. Algorithms are referenced by the section where they were described in this paper. Best performance is found towards the top-right corner (high precision and high recall).**

algorithms on the food types potato chips, carrot (representing food with wet crispy texture like lettuce) and apple. It seems that most of the algorithms evaluated by us could outperform the results of Amft et al. However, Amft et al. took into account the borders of automatically assigned chew event labels in their evaluation routine. For we used discrete time marks for chew events assigned by the algorithms, a direct comparison to the results of Amft et al. is not possible.

Evaluating records of the food types peanut, walnut and chocolate showed lower performances than the other types of food. This is seen in the texture properties of these food types. While potato chips, carrot and apple had crisp texture, peanut, walnut and chocolate were not as crispy as the former ones. The lack in crispness results in lower sound signal amplitudes in the records. Therefore, many chew events have not enough signal energy to exceed the threshold value and hence, are not detected by the algorithms. Different threshold values for food types of different texture may overcome this drawback. However, this investigation was not in the scope of this paper and will be considered in future research.

**Table 1. Performance values achieved from each algorithm on the test records. Algorithms are referenced by the section where they were described in this paper. Acc, Pre and Rec stand for accuracy, precision and recall, respectively. All values are given in percent.**

| algorithm | section 4.1 | | | section4.2 | | | section 4.3 | | | section 4.4 | | | section 4.5 | | | section 4.6 | | | section 4.7 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Acc | Pre | Rec | Acc | Pre | Rec | Acc | Pre | Rec | Acc | Pre | Rec | Acc | Pre | Rec | Acc | Pre | Rec | Acc | Pre | Rec |
| **food type** | | | | | | | | | | | | | | | | | | | | | |
| potato chip | 74 | 92 | 81 | 63 | 79 | 87 | 52 | 72 | 87 | 54 | 80 | 71 | 64 | 88 | 74 | 71 | 91 | 79 | 65 | 81 | 86 |
| peanut | 39 | 90 | 45 | 53 | 75 | 79 | 47 | 76 | 68 | 45 | 78 | 62 | 32 | 86 | 38 | 41 | 87 | 48 | 49 | 77 | 69 |
| walnut | 36 | 92 | 40 | 60 | 81 | 79 | 52 | 80 | 70 | 50 | 82 | 65 | 31 | 86 | 37 | 42 | 89 | 48 | 57 | 82 | 72 |
| carrot | 82 | 96 | 85 | 73 | 90 | 82 | 74 | 87 | 87 | 66 | 91 | 73 | 72 | 91 | 80 | 80 | 96 | 84 | 72 | 91 | 80 |
| apple | 78 | 94 | 84 | 70 | 84 | 86 | 65 | 80 | 86 | 64 | 87 | 75 | 67 | 88 | 78 | 77 | 93 | 83 | 68 | 84 | 83 |
| chocolate | 19 | 77 | 27 | 42 | 68 | 78 | 31 | 68 | 58 | 31 | 71 | 53 | 7 | 65 | 15 | 26 | 79 | 35 | 40 | 70 | 71 |
| **overall** | 57 | 91 | 64 | 54 | 75 | 81 | 52 | 76 | 76 | 50 | 79 | 67 | 48 | 86 | 57 | 56 | 88 | 65 | 53 | 77 | 76 |

Overall performance values of the maximum spectral band energy algorithm and the "BeatThis" algorithm rank best in precision and recall. Both algorithms are based on the analysis of different spectral bands of the signal. They are based on the high energy content in all components of the spectrum of chew event sounds. Algorithms based only on overall signal energy showed lower performances. Additionally, the maximum spectral band energy algorithm is less complex and needs less computing effort than the "BeatThis" algorithm. This algorithm is expected to be able of being integrated into wearable devices, which have limited resources due to limited battery lifetime. Nishimura and Kuroda achieved error rates of 1.93 % on average using the "chew-like" signal detection algorithm and verification using models based on MFCC features. They recorded chewing sounds of 5 different types of food. This remarkable result could not be reproduced in our investigations. The difference is seen in the way Nishimura and Kuroda [6] collected the data. Their recording device used occlusion of the ear to prevent recording of environmental sounds. We avoided ear occlusion and hence, recorded background noise of the recording room. Future research will show if signal enhancement algorithms can reduce the noise level and improve the chew event detection performance. Shuzo et al. did not give performance values of the counting of mastications by their algorithm [11].

However, all evaluations were carried out only on records of six types of food. Other types of food may be detected better or less well than the selected ones. To investigate the detection performance on a wide variety of different food types, studies on records of participants eating more different types of food are planned. These studies will show whether similar food properties in different types of food allow a reasonable grouping of the corresponding sound models. Doing so could also help to estimate and classify the properties like density, viscosity, liquidity or other textural properties of the consumed food. Additionally, the records contain mainly clean food intake sounds without background noise or environmental signals. We plan to carry out the chew event detection on records of different environmental sounds recorded by our sensor system. For this investigation, a pre-processing step containing food intake activity detection [7] or signal enhancement may reduce the influence of noisy recording conditions. This and optimizing the algorithms evaluated in this work will give a possibility for accurately detection of chew events in records of chewing sounds under free-living conditions. In future investigations, we want to create a really wireless sensor system with the ability of on-body data evaluation. For this purpose, we need to consider low-power storage and signal processing technologies for a battery-powered device. Furthermore, the usability of the device must be investigated in long-time studies under free-living conditions.

## 7. CONCLUSION

A timesaving and pervasive method to protocol human food intake behavior is required to fight the challenge of obesity. For automated analysis of human food intake patterns, detecting chew events plays an important role. This is not only for counting chew events but also for the estimation of the bite weight of consumed food. In this paper, we analyzed different algorithms for detection of chew events. Some of them were taken from literature of food intake analysis. Others came from beat detection in music. The algorithms were evaluated on a database recorded from fifty-one participants with our sensor system and annotated manually. Hence, a big variety of individual chewing styles was contained in the records. Especially algorithms, which detect signal energy maxima in different spectral bands, showed promising results. However, the algorithms need to be optimized to reduce the number of insertions and deletions. The algorithms will be carried out on recorded reference sounds in combination with food intake activity detection algorithms in future studies. This will show if chew event detection works in noisy environments. Overall, an important step to develop a robust method for food intake pattern analysis and bite weight estimation has been carried out.

## 9. REFERENCES

[1] Amft, O., Stäger, M., Lukowicz, P. and Tröster, G. 2005 Analysis of chewing sounds for dietary monitoring. In Proceedings of 7th International Conference on Ubiquitous Computing (Tokyo, Japan, September). UbiComp'05. 56–72.

[2] Amft, O., Kusserow, M. and Tröster, G. 2009. Bite weight prediction from acoustic recognition of chewing. *IEEE Trans. Bio.-Med. Eng.* 56, 6, 1663-1672.

[3] Amft, O. 2010. A wearable earpad sensor for chewing monitoring. In *Proceedings of IEEE Sensors 2010 Conference* (Waikoloa, Hawaii, USA, November 01-04, 2010). 222-227.

[4] Cheng, K., Nazer, B., Uppuluri, J. and Verret, R. *Beat This – A Beat Synchronization Project.* DOI= http://www.clear.rice.edu/elec301/Projects01/beat_sync/index.html.

[5] Drake, B. K. 1963. Food crushing sounds. an introductory study. *J. Food Sci.* 28, 233-241.

[6] Nishimura, J. and Kuroda, T. 2008. Eating habits monitoring using wireless wearable in-ear microphone. In *Proceedings of International Symposium on Wireless Pervasive Communications* (Santorini, Greece, May 07-09, 2008). ISWPC'08. 130-133.

[7] Sebastian Päßler, Matthias Wolff, Wolf-Joachim Fischer: Food intake monitoring: An acoustical approach to automated food intake activity detection and classification of consumed food. Physiol. Meas. 33 (2012) 1073-1093. (URL: http://stacks.iop.org/0967-3334/33/1073)

[8] Päßler, S., Wolff, M. and Fischer, W.-J. 2011. Chewing sound classification using a grammar based classification algorithm. In *Proceedings of Forum Acusticum 2011* (Aalborg, Denmark, June 26 – July 01, 2011). FA2012. 39-44.

[9] Scheirer, E. D. 1998. Tempo and beat analysis of acoustic musical signals. *J. Acoust. Soc. Am.* 103, 1, 588-601.

[10] Schoeller, D. A. 1995. Limitations in the assessment of dietary energy intake by self-report. *Metabolism. Res.* 44, 18–22.

[11] Shuzo, M., Komori, S., Takahashima, T., Lopez, G., Tatsuta, S., Yanagimoto, S., Warisawa, S., Dalaunay, J. and Yamada, I. 2010. Wearable eating habit sensing system using internal body sound. *J. of Advanced Mechanical Design, Systems and Manufacturing* 4, 1, 158-166.

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