Bite weight prediction from acoustic recognition of chewing

Bite Weight Prediction From Acoustic Recognition of Chewing

Oliver Amft*, *Member, IEEE*, Martin Kusserow, *Student Member, IEEE*, and Gerhard Tröster, *Senior Member, IEEE*

***Abstract*—Automatic dietary monitoring (ADM) offers new perspectives to reduce the self-reporting burden for participants in diet coaching programs. This paper presents an approach to predict weight of individual bites taken. We utilize a pattern recognition procedure to spot chewing cycles and food type in continuous data from an ear-pad chewing sound sensor. The recognized information is used to predict bite weight. We present our recognition procedure and demonstrate its operation on a set of three selected foods of different bite weights. Our evaluation is based on chewing sensor data of eight healthy study participants performing 504 habitual bites in total. The sound-based chewing recognition achieved recalls of 80% at 60%–70% precision. Food classification of chewing sequences resulted in an average accuracy of 94%. In total, 50 variables were derived from the chewing microstructure, and were analyzed for correlations between chewing behavior and bite weight. A subset of four variables was selected to predict bite weight using linear food-specific models. Mean weight prediction error was lowest for apples (19.4%) and largest for lettuce (31%) using the sound-based recognition. We conclude that bite weight prediction using acoustic chewing recordings is a feasible approach for solid foods, and should be further investigated.**

***Index Terms*—Algorithm implementation, biosignal processors, signal and image processing.**

I. Introduction

AUTOMATIC dietary monitoring (ADM) aims at simplifying the reporting of individual eating behavior for weight and diet coaching programs [1]. Moreover, it may become a vital tool for dietary supervision in clinical observation of obese patients and to support independent living of elderly individuals. ADM is based on ubiquitous on-body or ambient sensors and pattern recognition techniques to derive eating behavior information. It proposes a novel monitoring paradigm, compared to manually recording of dietary activities, as it is currently performed using daily self-reports [2]. For the success of ADM, it is essential that it provides similar information detail on eating behavior as derived through self-reports today.

Besides meal schedule and food type of each intake, self-reports typically record amount of consumed foods. Such food amount data provide essential information on balance of nutrient composition and portion size. While respondents could count food amount in some cases, such as the number of apples consumed over a day, many foods require weighting before the intake. However, weighting every food item adds a substantial burden for respondents to follow a natural lifestyle. This continuous manual monitoring effort hampers participant compliance in coaching programs [3]. Moreover, self-reports are biased due to misreporting of respondents. Reporting errors depend on various social and personal aspects [4], [5], and increase with monitoring duration. Typically, report entries are omitted, partially completed, or backfilled, as perception of desirable intake patterns evolve. An error variation between 50% under- and overestimation was observed in healthy individuals and patients [4], [6].

The goal of this paper was to evaluate the prediction of food weight in individual bites using an ear-pad chewing sound sensor. We refer to bite weight as a quantity of food amount that is ingested into the mouth with each bite taken. The prediction utilizes a sound-based recognition of chewing cycles that provides structural and timing variables of chewing sequences. In our approach, bite weight prediction models are selected based on recognized food types. As required for a robust ADM system operation, we adopted in our evaluation a habitual food consumption protocol, including unconstrained food size selection and freestyle chewing. We demonstrate that a robust food classification is feasible using this recognition procedure. Bite weight prediction based on the sound-based recognition was subsequently compared to a semisupervised detection–prediction scheme using muscle activity recorded from surface electromyography (EMG) electrodes.

A. Chewing Recognition for ADM

Most ADM approaches focus on activities, such as upper body and arm motion during intake [7], chewing [8], and swallowing [9]. An essential requirement for ADM systems is that deployed sensing solutions are ubiquitously integrated, protect privacy, and minimize interference with daily activities of their user. For example, chewing detection based on EMG, which is used for comparison in this paper, may not be acceptable since EMG electrodes are applied in visible facial regions. Moreover, skin preparation and precise positioning of electrodes are required to ensure acceptable signal quality.

A particularly interesting source of information for bite weight prediction is the chewing microstructure used to consume a bite. The chewing microstructure can be described as a sequence of chewing cycles (closing and reopening movements of the mandible) used to decompose food pieces from ingestion

Manuscript received July 2, 2008; revised December 3, 2008. First published March 4, 2009; current version published June 10, 2009. This work was supported by the EU MyHeart project and the Swiss State Secretariat for Education and Research. *Asterisk indicates corresponding author.*

*O. Amft is with Signal Processing Systems, Technische Universiteit (TU) Eindhoven, NL-5600 MB Eindhoven, The Netherlands, and also with the Wearable Computing Laboratory, ETH Zurich, CH-8092 Zurich, Switzerland (e-mail: amft@ieee.org).

M. Kusserow and G. Tröster are with the Wearable Computing Laboratory, Eidgenössische Technische Hochschule (ETH) Zurich, CH-8092 Zurich, Switzerland (e-mail: kusserow@ife.ee.ethz.ch; troester@ife.ee.ethz.ch).

Digital Object Identifier 10.1109/TBME.2009.2015873

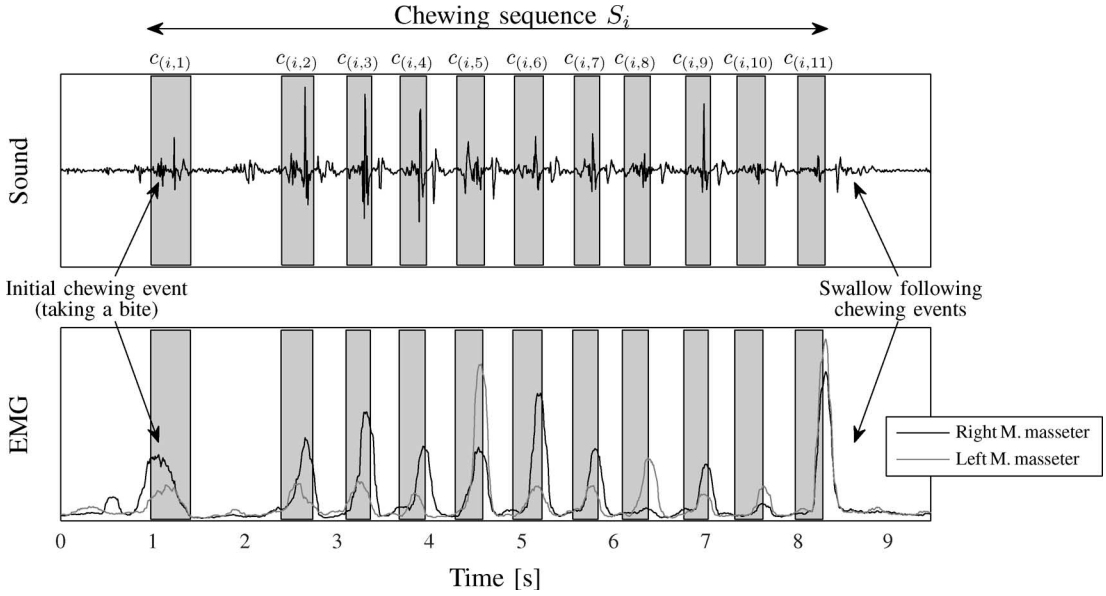

Fig. 1. Illustration of an apple chewing sequence S_i. Waveforms show data from the ear-pad chewing sound sensor (upper plot) and averaged rectified EMG electrodes (lower plot). The mandible closing phases of individual chewing cycles $c_{(i,j)}$ are marked as shaded areas.

into the mouth (bite) until swallowing [10]. Fig. 1 illustrates a chewing sequence for one ingested bite. We expected that microstructure variables of chewing sequences adapt according to the food weight in a bite.

Our previous investigations have shown that bone-conducted food breakdown sounds can be recorded by a miniature microphone at the ear canal [8]. Based on this sensing concept, food category and individual chewing cycles could be recognized from acoustic pattern models [11]. The current paper extends the robustness of food and chewing recognition using the chewing sound sensing approach.

B. Relation of the Chewing Microstructure to Bite Weight

Food ingested into the mouth excites different oral receptors that convey sensory information on material properties to the brain stem. Most important stimuli are food texture, such as crispness and hardness, size, shape, as well as flavor [12], [13]. Chewing continuously adapts to these stimuli, targeting an efficient food breakdown and creation of a food bolus that can be swallowed [14]. Intra-individually, this adaptation process is fairly constant. Using controlled settings and constant food stimuli, no significant differences were found in several chewing tests, when monitoring mandibular movement, muscle activity, and chewing microstructure [15], [16]. This stability of the microstructure is a key aspect to derive personalized bite weight prediction models in this paper.

A relation to chewing microstructure was previously observed for prescribed bite sizes from the same food [13]. Most consistent reports exist for variables measuring number of chewing cycles and chewing sequence duration from ingestion to swallowing. Both variables increased with bite size for artificial food [17] and three natural foods using fixed sizes [18].

Many investigations of the chewing microstructure analyzed masticatory performance in prescribed chewing assessments [19], [20]. However, the prediction of bite weight from chewing microstructure variables was neither investigated for fixed bite weights nor in habitual chewing. Hence, it is not clear how habitual bite weight of different natural foods is reflected in chewing behavior. Moreover, the chewing microstructure has not been recognized from continuous chewing sound data in previous works.

C. Food Selection for This Study

Food texture and material structure provides vital features for food discrimination from chewing sounds [8], [21]. We had previously confirmed the relation of sound patterns and food groups using an event recognition in specific texture types [11].

Food texture groups can be defined in various forms, as they relate to food perception and jargon of human panelists [22]. For the purpose of this paper, we refer to two texture groups and included three different foods: 1) *wet-crisp* structures from naturally grown foods, such as *apple* and *lettuce*, and 2) *dry-crisp* foods, such as *potato chips*. For our interest in ADM, the first group is particularly relevant, since the consumption of fruits and vegetables is vital for a balanced food selection [23].

The foods selected for this study served two purposes. First, they allowed us to evaluate a food-specific discrimination of similar sound emission groups (wet- and dry-crisp textures). Second, we chose them to study habitual bite size selection and bite weight prediction in different bite weight ranges. Regarding this second goal, foods were chosen to cue different weight selections, e.g., potato chips had typically the lowest bite weights, while apple bite weights were largest.

II. Experimental Procedure

A. Study Protocol

Eight volunteer students (two females and six males) aged between 20 to 35 years were recruited from different Eidgenössische Technische Hochschule (ETH) departments through advertisements. All participants had natural dentition, and no known history of chewing and swallowing abnormalities. Further exclusion criteria were disorders or audible sounds of the temporomandibular joints, as well as food allergies. A prerecording interview was conducted with each participant in the measurement room for familiarization. The recording procedure was explained; however, the specific goal of this investigation was not mentioned. Participants were subsequently invited for an individual recording session around midday.

Participants were asked to eat the following foods: potato chips (Chio chips "Ready salted", ∼25 pieces, total: 20 g), mixed lettuce (containing endive, sugar loaf, frisée, raddichio, chicory, and arugula, total: ∼55 g), and one apple ("Jangold", total: ∼110 g). The food weights indicate approximate values, since participants were allowed to eat as much as they liked from individual foods. They chewed and swallowed all foods in their habitual style. From potato chips, a few chips were taken for each bite with the hand, lettuce was consumed using fork and knife, and apples were eaten by taking bites from the skinned fruit. The apple core was not consumed.

All participants were familiar with the food types. None of them expressed a dislike or problems to chew or swallow the selected foods. Participants were allowed to move, drink water, and speak during recording sessions. Session duration was not constraint, since participants were eating/drinking at their individual pace. Informed consent was obtained from each participant. The study protocol was reviewed and approved by the ETH ethics committee.

B. Data Recording

Chewing sound was recorded using a miniature microphone (Knowles, TM-24546) embedded in a custom ear-pad device. Ear occlusion of this pad was kept low, in a way that participants could hear room-level conversation at the applied side. While low occlusion decreases signal-to-noise ratio, it is needed for wearer comfort and safety reasons. In contrast, high occlusion prevents air circulation to the ear canal, and disturbs sense of balance. A low noise level was maintained in the recording room, similar to a restaurant environment.

The sound signal was constantly amplified (M-Audio, Audio Buddy) and sampled at 44 kHz, 16 bit using a computer-connected sound device (ESI, U46DJ). We expect that this sampling rate could be reduced without impact on performance. However, optimizing sampling rate was not a goal of this paper. Surface EMG was recorded bilaterally from M. masseter at 2 kHz, 24 bit, and bandpass-filtered in the recording device (MindMedia, Nexus 10). The plate weight was recorded at ∼1 Hz with a resolution of 0.1 g. We used a computer-connected weight scale (Kern, 572-45) that was embedded into a custom-build table. The scale performed an automatic measurement stabilization. Synchronization marks were embedded in data of all sensor modalities during recordings. These marks were subsequently used to align data streams after recording each session. We estimate that the remaining alignment jitter due to our recording setup was below 0.1 s.

An observer controlled the data recording during each session, and annotated chewing sequences and swallowing events. In a postrecording step, all annotated sequences were reviewed, start/end times adapted, and swallowing events marked for exclusion by inspecting signal waveforms. A total dataset of 8.64 h was recorded and annotated. The average length per participant was 64.83 min (standard deviation 14.6 min).

C. Chewing Annotation

In order to train and validate food pattern models, an annotation of chewing cycles in all recorded data was required. This annotation consisted of location marks for individual chewing events (mandible closing phase of a chewing cycle) and food type information for each such event. We will further denote this annotated phase as chewing event to differentiate it from a complete chewing cycle.

All annotations of chewing events were performed manually in a postrecording step, by reviewing sound and EMG waveforms as well as listening to chewing sounds. As our goal was to obtain an accurate and consistent evaluation baseline, all chewing annotations were performed by one observer. This method is precise in identifying every chewing event before the main food bolus is swallowed. However, it is expensive and time-consuming for large chewing data sets, as in our study. To reduce annotation efforts, we tested a sound-energy-based segmentation of chewing events. However, this segmentation failed to correctly identify events, since sound energy is highly variable during chewing. Energy during mandible reopening may even exceed sound energy during closing phases. Although less pronounced, we observed similar issues for a chewing segmentation using the EMG amplitude. For performance comparison to our manual annotation, we included a semisupervised EMG-based chewing detection in the evaluation (Section V-A).

Fig. 1 shows sample waveforms from the sound sensor and EMG electrodes for a chewing sequence of one apple bite. While intermediate swallowing can occur, none was observed in the depicted example. The mandible closing phases of each cycle are marked as shaded areas. In total, 7910 chewing cycles were identified and annotated in 504 chewing sequences.

III. Chewing Recognition Procedure

We developed a recognition procedure to derive chewing events from the sound sensor data. This procedure attempts to identify temporal bounds for each individual chewing event, and classifies events regarding their food type. In order to identify chewing events a feature similarity search (FSS) was applied for each food and participant. Subsequently, a food type classification was used to determine food type for each detected chewing event. A sequence voting was used to determine food type of each chewing sequence. Events detected by all food-specific FSS instances were fused by comparing concurrently

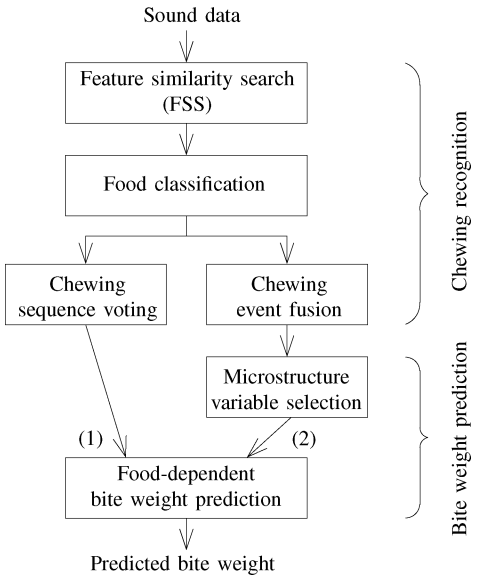

Fig. 2. Methodology to recognize chewing events and predict bite weight. Bite weight prediction uses: (a) food type and (b) chewing microstructure information. The procedure for *chewing recognition* is detailed in Section III. The method for *bite weight prediction* is presented in Section IV.

occurring events. This event fusion step retained the events with the highest classification model confidence.

Fig. 2 illustrates the entire evaluation procedure to identify chewing events, classify food type, and predict bite weight. In this section, the recognition procedure is presented in detail. All evaluation steps regarding the prediction of bite weights are discussed in Section IV.

A. Feature Similarity Search (FSS)

FSS is an online event recognition algorithm, based on a variable-length feature pattern search. The algorithm has been introduced in our previous works, e.g., [11]. We used FSS here to recognize temporal boundaries of chewing events in sound data. FSS is particularly applicable for this task, since the sound data may contain other arbitrary noises (NULL class) besides chewing events.

FSS utilizes a data description approach to model chewing events from food-specific sound features, and discriminate these events from noise. We used a training set $\mathcal{F}^+$ of feature vectors $\mathbf{f} = (f_1, \ldots, f_{N_F})$ to describe the pattern of a chewing event: $c_{(i,j)} \rightarrow \mathbf{f}(t_{(i,j)}, l_{(i,j)})$. Fig. 1 illustrates this concept. Here, $c_{(i,j)}$ is the jth chewing event in chewing sequence $\mathcal{S}_i$, $\mathcal{S}_i = c_{(i,1)}, \ldots, c_{(i,M_i)}\}$, with $\mathcal{S}_i \in \mathcal{S}$ and total number of events M_i. Each event $c_{(i,j)}$ has two temporal parameters, time of occurrence $t_{(i,j)}$ and duration $l_{(i,j)}$.

Using the event model, FSS then searches through sound data. A normalized Euclidean distance function d was used to evaluate feature vector $\mathbf{f}(t, l)$ at position t and potential duration l. Hence, for each t, a distance is obtained, measuring the similarity of a data section with duration l to a chewing event model. All paramters for d as well as search bounds for l were estimated using training set data.

A distance threshold D_{Thres} was derived for each FSS model to omit NULL class data sections. D_{Thres} was determined by evaluating the recognition performance on training set data, containing chewing events ($\mathcal{F}^+$) and disjunct NULL class data ($\mathcal{F}^-$). We targeted a high sensitivity in order to retrieve at least 90% of all training set chewing events.

We selected a constant interval of $\Delta t = 125$ ms for each evaluation of d. While this interval limits the temporal resolution of retrieved chewing events, it reduces processing requirements compared to an evaluation for every sampled data point. The interval was acceptable for bite weight prediction, since temporal information at a lower resolution was not expected.

1) Dataset Cross-Validation: To derive robust recognition performance results from our dataset, we applied a ten-fold cross-validation to select training and validation data for FSS, feature selection, and all subsequent recognition steps. The dataset was partitioned into ten sections, while maintaining complete chewing sequences in each section. For each iteration, nine data sections were used for training and one for validation. Hence, each section was used once for validation.

2) Derived Feature Set: A total of 264 sound features were extracted from sound data. This set had been used in earlier studies [11]. It consists of the following feature subsets: log-band spectral energy, cepstral coefficients, and linear predictive coefficients (ten features each). Moreover, skewness, kurtosis, and tristimulus (total: 33 features) were included. All spectral features were computed from a 512-point fast Fourier transform (FFT). All of these features had been devised for audio analysis before [24]. In addition, we used a logarithmic-sized bin distribution (maximum frequency: 22.05 kHz) to derive spectral energy bands. Previous chewing sound studies had observed spectra up to 8 kHz; however, components below 3 kHz were most important to describe foods [21]. The log-band spectral energy subset reflect these observations, as it uses small bins for low frequencies.

For all features, we computed mean and variance using a sliding window of 512 samples without overlap. We derived features for the entire chewing event as well as three evenly divided partitions of each event. These partition features were included to capture temporal patterns of chewing events in a spatial feature representation.

3) Feature Selection: To select an adapted feature subset, we deployed a feature relevance and independence filtering using training set data. While we did not confirm that this procedure is an optimal strategy among selection methods, it required only small adaptations to work with the FSS data description. For FSS, correct events and NULL class have a large skew; consequently, any selection algorithm should not consider a class prior.

In the first step, relevance w_R of all features elements $f_n \in \mathbf{f}$ was determined. For this purpose, we computed the absolute differences between training set feature distributions in correct events ($\mathcal{F}^+$) and NULL class ($\mathcal{F}^-$). Distributions were estimated using histograms of bin size $N^{(1/3)}$, where N is the number of training feature vectors.

The second step refines feature relevance ranking by estimating feature independence, introduced in [25]. This step aims to

select a relevant feature subset, while minimizing redundancy. Independence I was determined from correlations among features, $f_n, f_m \in \mathbf{f}$ in training set events $\mathcal{F}^+$, using Spearman's correlation coefficient $\rho(f_n, f_m)$ [26]

$$I(f_n, f_m) = \sqrt{1 - \rho(f_n, f_m)^2} \qquad m \neq n\,. \tag{1}$$

Subsequently, an iterative scheme was used, starting with the highest relevance-weighted feature. In each iteration i, an additional feature was selected that yielded the highest combined weight w_C, when evaluated against all previously selected features $f_{h_k}, 1 \leq h_k \leq N_F$ (2). In total, 20 features were selected for the FSS step

$$w_C(f_n) = \sum_{k=1}^{i} w_R(f_n) I\big(f_n, f_{h_k}\big), \qquad h_k \neq n\,. \tag{2}$$

In the evaluation, we observed that cepstral coefficients and log-band spectral energy from low-frequency bands were selected most often. We also tested a relevance selection based on Mann–Withney–Wilcoxen [27] without further performance improvements. Wrapper-based approaches were not considered here, since repeated FSS training becomes very processing intensive for large datasets.

B. Food Classification

The FSS step was used to determine chewing event bounds for individual foods. As foods had similar textures, this lead to between-food confusions of FSS instances, and may, subsequently, select a wrong bite weight prediction model. We applied an additional food classification in this step for all chewing events. A nearest centroid classifier was trained based on a Fisher's linear discriminant feature transformation [28]. The same dataset cross-validation (partitioning of training and validation sets) was reused. We implemented the classification using all features derived for FSS, before feature selection. We computed a confidence for each chewing event by normalizing classifier distances with the largest class distance.

C. Chewing Event Fusion and Sequence Voting

Temporal event overlaps, as a result of independent FSS instances for each food, were pruned after the classification using an event comparison. This method filters all event detection results and retains those, which obtained largest food classification confidences. All chewing events obtained in this step were subsequently used for analyzing chewing microstructure.

The final food type was determined from a majority vote among all classified events in one chewing sequence. This approach is reasonable, since food type will not change within one chewing sequence. The food type result was used to select bite weight prediction models.

IV. Bite Weight Prediction Method

Based on recognized chewing events, we extracted variables to describe the chewing microstructure. Using both, microstructure variables and identified food type, we subsequently predicted bite weight.

TABLE I
MICROSTRUCTURE VARIABLES COMPUTED FOR EACH CHEWING SEQUENCE $\mathcal{S}_i$ AND SECTIONS WITHIN $\mathcal{S}_i$ (SEE SECTION IV-A)

Pos	Description	Law for sequence $\mathcal{S}_i$ $\forall i : 1 \leq j \leq \hat{M}_i$ *
1	Nr of chewing events	$\hat{M}_i$
2	Total chewing duration	$t(i,\hat{M}_i) - t(i,1)$
3	Mean event duration $\bar{l}_i$	$\sum_j l(i,j)/\hat{M}_i$
4	Variance of event duration	$\sum_j \big(l(i,j) - \bar{l}_i\big)^2/\hat{M}_i$
5	Slope of event duration	$cov\big(l(i,j), j\big)/var(j)$
6	Chewing speed	$\hat{M}_i/\big(t(i,\hat{M}_i) - t(i,1)\big)$
7	Slope of chewing speed	$cov\big(1/l(i,j), j\big)/var(j)$
8	Mean signal energy $\bar{E}_i$	$\sum_j E(c(i,j))$

*) Iterator j denotes all chewing events within $\mathcal{S}_i$ or section of $\mathcal{S}_i$.

A. Microstructure Variables and Relevance Analysis

Eight base variables were defined, as summarized in Table I. This base set was computed from each entire chewing sequence, three evenly partitioned temporal sections, as well as first five, and first three chewing events only (total: 50 variables). In addition, we computed event duration and mean signal energy from each first chewing event in a sequence. These variables, computed for sections within chewing sequences, allowed us to evaluate whether dynamic changes in the microstructure are relevant for bite weight prediction.

The correlation of each variable v_n with bite weight W was analyzed using Spearman's correlation coefficient ρ. Correlation results were summarized in relevance $w_V(v_n)$ for all participants

$$w_V(v_n) = \sum_{\text{Participants}} \big|\, \rho(v_n, W) \,\big|\,. \tag{3}$$

B. Prediction Model

We deployed a multiple linear regression model of the form

$$\hat{W}_i = a_0 + \sum_{k=1}^{N_V} a_k v_{ik} \tag{4}$$

for bite weight prediction. The microstructure variables are represented by $v_1, \ldots, v_{N_V}$, where N_V is the total number of variables in a prediction model. Result $\hat{W}_i$ denotes the predicted bite weight for a particular chewing sequence $\mathcal{S}_i$. Food-specific coefficients $a_0, a_1, \ldots, a_{N_V}$ were found by a least-squares fit on training data. The weight prediction was performed by a leave-one-out analysis to estimate prediction errors.

We investigated a stepwise regression fit to select informative variables among all microstructure variables. However, we observed a similar performance compared to a manual preselection based on variable relevance analysis (Section IV-A before).

V. Results

A. Recognition of Chewing Events

Recognition performance of chewing events was analyzed using Precision and Recall. These metrics are frequently used to assess event spotting performance (insertion and deletion

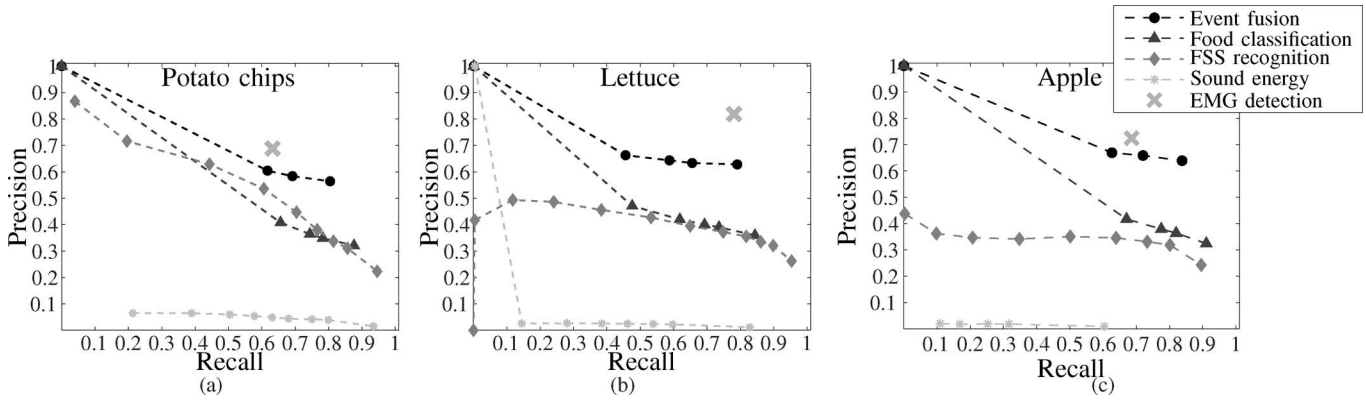

Fig. 3. Food-specific performance of the recognition steps in validation data (confidence threshold sweeps according to Section V-A) in comparison to a sound energy-based detection approach (*sound energy*). *EMG detection* shows the performance of a semisupervised EMG-based chewing detection in annotated chewing sequences. Best performance is found towards the top-right corner (high precision and high recall).

errors). A detailed description can be found in [7]. To account for jitter between recognized and annotated event bounds, a soft-alignment procedure was applied [11]. Events with a boundary deviation below 50%, with respect to the annotated event duration, were scored as correctly recognized.

All recognition steps: FSS, food classification, and event fusion were evaluated in validation data using the cross-validation procedure (see Section III-A.1). For comparison, we performed a sound energy-based detection using the FSS algorithm with signal energy as single feature. Fig. 3 shows food-specific performance visualizations, averaged for all participants. These visualizations were obtained by testing the validation set performance for different confidence thresholds in retrieved chewing events. The graphs show a line of best-performance points obtained for each recognition step.

These results indicate a good performance of our recognition procedure, despite very similar sound patterns in the selected foods. In particular, the classification and event fusion helped to refine recognition results. Overall a recall of 80% with a precision at 60%–70% was achieved for each food. As expected, the sound energy-based detection did not achieve a practically useful precision. This result can be attributed to natural variability in chewing sound energy and arbitrary noise in the dataset.

B. Semisupervised EMG Detection

Fig. 3 additionally shows performance results for a semisupervised EMG-based chewing detection. The detection was applied in annotated bounds of chewing sequences only. It was implemented according to chewing behavior investigations [20]. In this implementation, the rectified EMG signal was mean-filtered with a sliding window of 125 ms. A threshold was used to derive chewing events, and set to the signal level before chewing onset plus 1 standard deviation of the signal in each processed chewing sequence. Our evaluations showed that this EMG detection misses many events toward the end of sequences, presumably due to reduced muscle contractions. As a consequence, this EMG-based detection clearly underperformed the manual annotation of chewing events.

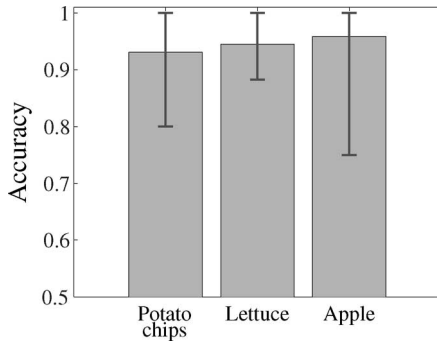

Fig. 4. Class-normalized accuracy of chewing sequence majority voting. Min.–Max. values show participant-specific result variation.

C. Food Classification

Fig. 4 shows the final food classification performance for chewing sequences, using event majority voting. Events used for this voting were obtained from the food classification step at a recall of ∼80% for all foods. In this evaluation, we used the class-normalized accuracy $a = \frac{1}{N_C} \sum_{c=1}^{N_C} \frac{\mathcal{S}\ \text{Recognized}_c}{\mathcal{S}\ \text{Relevant}_c}$ as performance measure. Here, N_C is the total number of classes, and $\mathcal{S}$ Recognized$_c$ and $\mathcal{S}$ Relevant$_c$ are the number of correctly identified and total chewing sequences from food type c, respectively. Any class skew was removed from the training set instances. For this final food classification, we obtained an average accuracy of 94%. This excellent performance supports our approach to select bite weight models based on food classification results.

D. Microstructure Relevance Analysis

Correlations of microstructure variables with bite weight were analyzed [using (3)] to determine commonly relevant variables for all participants. Fig. 5 shows a variable relevance map based on events retrieved from sound-based recognition and semisupervised EMG detection. High relevance ($w_V \geq 0.6$), hence large individual correlations, were observed for the variables number of chewing events and chewing duration, except for potato chips. We observed largest participant-specific

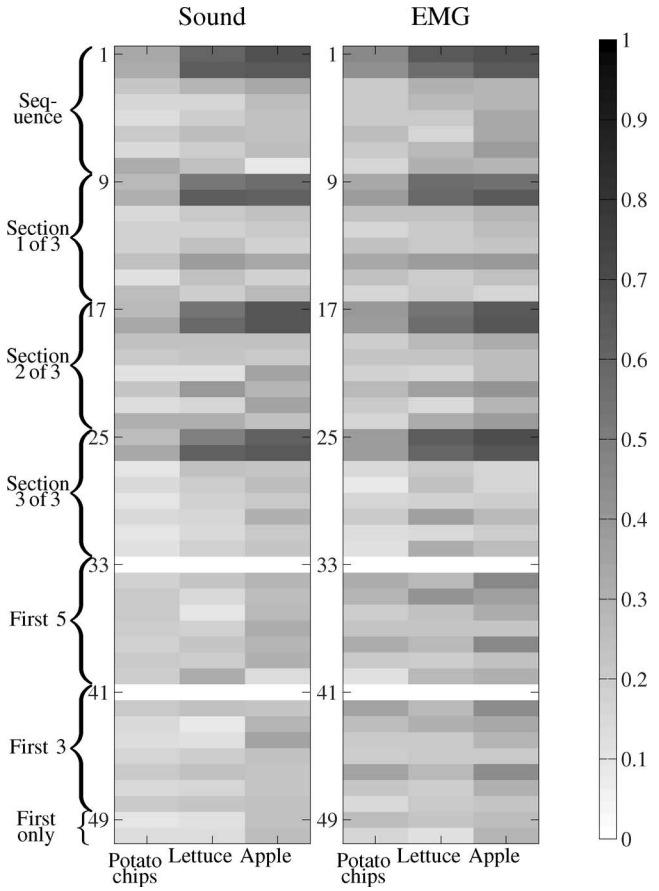

Fig. 5. Microstructure variable relevance (3) for sound-based recognition (left) and EMG detection (right). All variables are detailed in Section IV-A.

correlations for apple (up to 0.96). Overall, relevant variables were the same for sound-based recognition and EMG detection.

E. Bite Weight Prediction

Bite weight was predicted using a subset of four variables, which we determined from relevance analysis and stepwise regression fit. This subset contained number of chewing events and chewing duration for the entire sequence and section 1 of 3. The relevance analysis confirmed that that these variables were the most relevant ones for all foods.

Table II shows bite weight prediction errors for sound-based recognition and EMG detection. Moreover, we computed errors for an interindividual prediction model using the sound-based recognition and for a constant weight (average weight from second and third chewing sequence). The leave-one-out verification scheme was used for all results, except the constant weight prediction.

A constant weight prediction assumes that bite weights do not change for a specific food. We assumed that the second and third chewing sequences represented a good average weight. In our evaluation, this constant weight prediction marks a performance baseline. However, our results demonstrate that predictions using the chewing microstructure outperform this baseline for all foods.

TABLE II
PERFORMANCE OF DIFFERENT BITE WEIGHT PREDICTION APPROACHES

Metric	**Foods**		
Mean (SD)	**Potato chips**	**Lettuce**	**Apple**
Bite weight W [g]	0.8 (0.2)	2.3 (0.8)	7.8 (1.5)
Chews/Seq. (M_i)	26.9 (4.4)	20.0 (3.3)	14.9 (2.9)
	Sound-based recognition		
Absolute error [g]	0.2 (0.1)	0.6 (0.2)	1.4 (0.4)
Relative error [%]	**27.7 (9.5)**	**31.0 (5.5)**	**19.4 (4.3)**
	EMG detection		
Absolute error [g]	0.2 (0.1)	0.6 (0.2)	1.9 (1.1)
Relative error [%]	**26.5 (9.0)**	**28.9 (4.0)**	**27.8 (14.6)**
	Sound-based rec. (inter-individual)		
Absolute error [g]	0.2 (0.2)	0.8 (0.6)	2.3 (1.8)
Relative error [%]	**31.7 (30.6)**	**40.2 (38.2)**	**37.2 (37.1)**
	Constant weight[a]		
Absolute error [g]	0.3 (0.1)	0.9 (0.3)	3.3 (1.8)
Relative error [%]	**41.1 (25.8)**	**50.5 (29.8)**	**62.2 (33.8)**

[a]Average weight of 2nd and 3rd chewing sequence.

Overall lowest errors were achieved for the sound-based weight prediction of apple. Here, the average error was only 19.4%. This result demonstrates the applicability of a sound-based approach, compared to constant weight (error: 62%) and EMG-based predictions (error: 28%).

Fig. 6 illustrates cumulative intake curves that are frequently used to assess food-weight-related intake behavior [10]. We deploy them here to visualize the variability in bite weights, and qualitatively compare prediction results to actual values. The graphs show that our sound-based prediction closely follows the actually consumed food weight. A constant weight could predict intakes of low weight variations only, such as for potato chips in our study.

VI. DISCUSSION

The focus of this paper was to analyze the prediction of bite weights, being the smallest granularity of food intake. Our results demonstrate that even foods with very similar acoustic emissions could be recognized in continuous sound data using our recognition procedure. Moreover, with the dataset cross-validation, we obtained robust performance results.

A specific challenge faced in chewing recognition is the expensive manual data annotation to obtain a ground truth for evaluations. We tested alternative methods such as sound energy and EMG-based chewing detections to reduce this effort. Results for a EMG-based chewing detection showed that its performance is approximately 20% below our manual annotation, even in a semisupervised mode. Our observation of EMG detection errors is in agreement with reports from other studies, e.g., [29]. Further investigations are needed to refine semisupervised annotation techniques, potentially by combining EMG and sound data.

A consequence of this annotation challenge is a low number of foods that could be evaluated in our present study. Nevertheless, we chose these foods carefully, following our interest to monitor fruit and vegetable consumption, and studying the discrimination of similar acoustic textures. Moreover, the selected

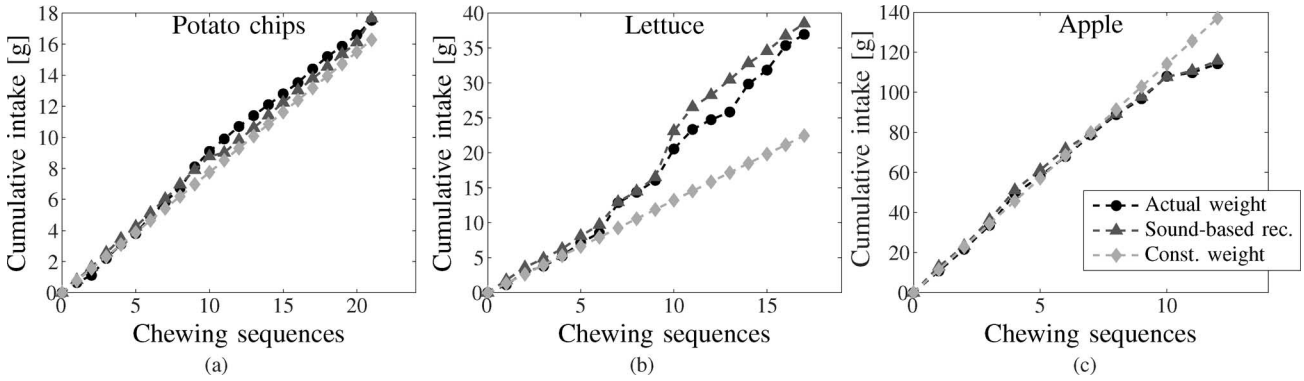

Fig. 6. Example cumulative intake curves illustrating bite weight prediction for sound-based recognition and constant weight approaches.

foods cover a spectrum of typical bite weights and consumption styles (see mean bite weights in Table II) relating to bite weight prediction, which was the focus of this paper. We expect that our prediction approach using a sound-based recognition will be feasible for other individually consumed solid foods. First, ADM systems that use the bite weight prediction may focus on a small set of relevant foods, e.g,. certain fruits and vegetables. Weight prediction models for new foods may even be derived from weight information initially provided by the user. To this end, we confirmed in a recent chewing study with three participants that even 19 solid foods could be classified from chewing sounds [1].

Our analysis of microstructure variables show that participants adapted their chewing behavior to the foods in a similar way. We observed consistent correlations of 0.7–0.96 with bite weight for several variables. Moreover, the same variables were the most relevant for all foods. Similar correlations to bite size were previously observed for artificial or fixed-sized foods only [18]. Our results for natural foods and habitual bite selection are promising, in particular, for the bite weight prediction approach presented in this paper.

The results also show that no microstructure adaptations within chewing sequences occur that could be captured by linear models. Further work should address whether other prediction concepts, potentially models of higher orders, could provide additional benefits.

Both, correlation and weight prediction results show that bite weight is not equally reflected in the chewing microstructure of all foods. Especially for foods with typically low bite weights, such as lettuce and potato chips, prediction errors are higher ($\sim$30%) than in apples (error: 19.4%). We concluded that chewing behavior does not adapt to these bite weights as it does for larger weights. A similar observation was made by a recent study on gum chewing of different weights [30]. There, a 1 g gum bolus resulted in the largest within-subject variability, suggesting that oral receptors are less sensitive to these low bite weight stimuli.

Nevertheless, for all foods in this investigation, a sound-based recognition of the chewing microstructure outperformed all other prediction approaches. Notably, this approach performed even better compared to a prediction using semisupervised EMG-based chewing detection. The prediction performance for apples even approaches the weight variation in this fruit type itself. Hence, the prediction error is as low as in accurately maintained self-reports, that record amount, instead of requiring food weighting. When compared to a "typical" performance of self-reports of up to 50% over- and underreporting [4], [6], our approach indicates clear advantages that are very promising. However, further investigations are required, that include additional foods and consumption patterns.

While altering lubrication of foods, such as buttering toast, lowers the total number of chews, the change in absolute numbers remained low [31]. Deterioration, e.g., in apples, may add an uncertainty in mass density of up to $\sim$10% [32]. Hence, both aspects may limit overall bite weight prediction accuracy, even for systems covering fruit and vegetables only. However, this additional error will still be acceptable, compared to errors and efforts for self-reports. At worst, if preparation or deterioration would modify the material structure, an acoustic food recognition may reject a food category. In this situation, a food recognition system may offer a selection of most likely foods and toppings. Once a user selection was made, the corresponding food model could be used for amount prediction.

VII. Conclusion

Techniques to automatically monitor eating behavior are very promising replacements for self-reporting assessments. However, it is essential for ADM to provide eating behavior information types as self-reports do today. This paper introduced a novel monitoring dimension to ADM: the prediction of food weight based on acoustic recognition of chewing.

Our study explored the relation between chewing microstructure and weight of individual bites taken. In particular, we considered a set of natural foods and habitual consumption patterns. As single modality, an ear-pad chewing sound sensor was used to recognize chewing cycles in continuous sound data. For individual chewing events, we achieved good recognition results (recall: $\sim$80% and precision: $\sim$60%–70%) and an excellent food classification performance (accuracy: $\sim$94%). These results support our approach to use the recognition procedure for selecting food-specific bite weight prediction models.

The weight prediction results showed that assessing bite weights from chewing microstructure information is a feasible approach. We confirmed that a constant bite weight

assumption fails with bite weight prediction errors of 60%, while our prediction using sound-based recognition achieved lower errors, down to 19.4% for apples. This result marks an important achievement in the development of alternative techniques to replace self-reports.

We expect that our approach can predict weights of other solid foods as well. Our future work will address the ambulatory evaluation of bite weight prediction and new annotation concepts for chewing.

ACKNOWLEDGMENT

The authors would like to thank all volunteers who participated in the food intake study. Moreover, the authors thank P. Colombani (Nutrition Biology Group of ETH Zurich) for his support during preparation of this study.

REFERENCES

[1] O. Amft and G. Tröster, "Automatic dietary monitoring: On-body sensing solutions for eating behavior monitoring," *IEEE Pervasive Comput.*, vol. 8, no. 2, pp. 62–70, Apr.–Jun. 2009.

[2] J. C. Witschi, "Short-term dietary recall and recording methods," in *Nutritional Epidemiology*, vol. 4, W. Willett, Ed., London, U.K.: Oxford Univ. Press, 1990, pp. 52–68

[3] P. M. O'Neil, "Assessing dietary intake in the management of obesity," *Obesity Res.*, vol. 9, no. 5, pp. 361S–366S, Dec. 2001.

[4] R. J. Hill and P. S. Davies, "The validity of self-reported energy intake as determined using the doubly labelled water technique," *Brit. J. Nutr.*, vol. 85, no. 4, pp. 415–430, Apr. 2001.

[5] K. R. Westerterp and A. H. C. Goris, "Validity of the assessment of dietary intake: Problems of misreporting," *Current Opin. Clin. Nutr. Metab. Care*, vol. 5, no. 5, pp. 489–493, Sep. 2002.

[6] D. A. Schoeller, "Limitations in the assessment of dietary energy intake by self-report," *Metab.: Clin. Exp.*, vol. 44, no. 2, pp. 18–22, Feb. 1995.

[7] H. Junker, O. Amft, P. Lukowicz, and G. Tröster, "Gesture spotting with body-worn inertial sensors to detect user activities," *Pattern Recognit.*, vol. 41, no. 6, pp. 2010–2024, Jun. 2008.

[8] O. Amft, M. Stäger, P. Lukowicz, and G. Tröster, "Analysis of chewing sounds for dietary monitoring," in *UbiComp 2005: Proc. 7th Int. Conf. Ubiquitous Comput.*(Lecture Notes in Computer Science), vol. 3660, M. Beigl, S. Intille, J. Rekimoto, and H. Tokuda, Eds. Heidelberg, Berlin, Germany: Springer-Verlag, Sep. 2005, pp. 56–72.

[9] O. Amft and G. Tröster, "Methods for detection and classification of normal swallowing from muscle activation and sound," in *PHC 2006: Proc. 1st Int. Conf. Pervasive Comput. Technol. Healthcare*, Piscataway, NJ: ICST, Nov. 2006, pp. 1–10. Doi: 10.1109/PCTHEALTH.2006.361624

[10] H. R. Kissileff and J. L. Guss, "Microstructure of eating behavior in humans," *Appetite*, vol. 36, no. 1, pp. 70–78, Feb. 2001.

[11] O. Amft and G. Tröster, "Recognition of dietary activity events using on-body sensors," *Artif. Intell. Med.*, vol. 42, no. 2, pp. 121–136, Feb. 2008.

[12] P. J. Lillford, "The materials science of eating and food breakdown," *MRS Bull.*, vol. 25, no. 12, pp. 38–43, Dec. 2000.

[13] A. Woda, K. Foster, A. Mishellany, and M. A. Peyron, "Adaptation of healthy mastication to factors pertaining to the individual or to the food," *Physiol. Behav.*, vol. 89, no. 1, pp. 28–35, Aug. 2006.

[14] R. Orchardson and S. W. Cadden, "Mastication," *The Scientific Basis of Eating* (Front Oral Biology Basel). Switzerland: Karger, 1998, vol. 9, pp. 76–121.

[15] W. E. Brown, M. Shearn, and H. J. H. Macfie, "Method to investigate differences in chewing behaviour in humans. I. Use of electromyography during chewing to assess chewing behavior," *J. Texture Stud.*, vol. 25, no. 1, pp. 17–31, Mar. 1994.

[16] C. Lassauzay, M. A. Peyron, E. Albuisson, E. Dransfield, and A. Woda, "Variability of the masticatory process during chewing of elastic model foods," *Eur. J. Oral Sci.*, vol. 108, no. 6, pp. 484–492, Dec. 2000.

[17] P. H. Buschang, G. S. Throckmorton, K. H. Travers, and G. Johnson, "The effects of bolus size and chewing rate on masticatory performance with artificial test foods," *J. Oral Rehabil.*, vol. 24, no. 7, pp. 522–526, Jul. 1997.

[18] F. A. Fontijn-Tekamp, A. van der Bilt, J. H. Abbink, and F. Bosman, "Swallowing threshold and masticatory performance in dentate adults," *Physiol. Behav.*, vol. 83, no. 3, pp. 431–436, Dec. 2004.

[19] M. A. Peyron, K. Maskawi, A. Woda, R. Tanguay, and J. P. Lund, "Effects of food texture and sample thickness on mandibular movement and hardness assessment during biting in man," *J. Dental Res.*, vol. 76, no. 3, pp. 789–795, Mar. 1997.

[20] R. González, I. Montoya, J. Benedito, and A. Rey, "Variables influencing chewing electromyography response in food texture evaluation," *Food Rev. Int.*, vol. 20, no. 1, pp. 17–32, Mar. 2004.

[21] N. DeBelie, M. Sivertsvik, and J. DeBaerdemaeker, "Differences in chewing sounds of dry-crisp snacks by multivariate data analysis," *J. Sound Vibration*, vol. 266, no. 3, pp. 625–643, Sep. 2003.

[22] A. Giboreau, C. Dacremont, C. Egoroff, S. Guerrand, I. Urdapilleta, D. Candel, and D. Dubois, "Defining sensory descriptors: Towards writing guidelines based on terminology," *Food Qual. Preference*, vol. 18, no. 2, pp. 265–274, Mar. 2007.

[23] B. J. Rolls, A. Drewnowski, and J. H. Ledikwe, "Changing the energy density of the diet as a strategy for weight management," *J. Amer. Dietetic Assoc.*, vol. 105, no. 5 (Suppl. 1), pp. S98–103, May 2005.

[24] G. Peeters, "A large set of audio features for sound description (similarity and classification) in the cuidado project,", Ircam, France, Tech. Rep., Apr. 2004.

[25] Q. Xu, M. Kamel, and M. M. A. Salama, "Significance test for feature subset selection on image recognition," in *Int. Conf. Image Anal. Recognit.* (Lecture Notes in Computer Science), vol. 3211, New York: Springer-Verlag, 2004, pp. 244–252.

[26] E. L. Lehmann, *Nonparametrics: Statistical Methods Based on Ranks, Revised*. Englewood Cliffs, NJ: Prentice-Hall, 1998.

[27] W. J. Conover, *Practical Nonparametric Statistics*, 3rd ed. New York: Wiley, 1998.

[28] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley Interscience, 2000.

[29] E. K. Kemsley, M. Defernez, J. C. Sprunt, and A. C. Smith, "Electromyographic responses to prescribed mastication," *J. Electromyogr. Kinesol.*, vol. 13, no. 2, pp. 197–207, Apr. 2003.

[30] A. M. Wintergerst, G. S. Throckmorton, and P. H. Buschang, "Effects of bolus size and hardness on within-subject variability of chewing cycle kinematics," *Arch. Oral Biol.*, vol. 53, no. 4, pp. 369–375, Apr. 2008.

[31] L. Engelen, A. Fontijn-Tekamp, and A. van der Bilt, "The influence of product and oral characteristics on swallowing," *Arch. Oral Biol.*, vol. 50, no. 8, pp. 739–746, Aug. 2005.

[32] D. Mitropoulos and G. Lambrinos, ""Delicious pilafa" apple density changes as a quality index of mass loss degradation during storage," *J. Food Qual.*, vol. 30, no. 4, pp. 527–537, Aug. 2007.

Oliver Amft (S'08–M'09) received the M.Sc. degree from Chemnitz Technical University, Chemnitz, Germany, in 1999, and the Dr. sc. ETH (Ph.D.) degree from ETH Zurich, Zurich, Switzerland, in 2008.

He is currently with ABB, Inc., Zurich, where he was a Senior Development Engineer and the R&D Project Manager from 2002 to 2004, and was engaged in product development of embedded communication systems, where since 2004, has been a Technology Consultant, and supports early development stages. He is currently an Assistant Professor at Technische Universiteit (TU) Eindhoven, Eindhoven, The Netherlands. He is also a Senior Research Advisor at the Wearable Computing Laboratory, ETH Zurich. He is interested in pervasive healthcare and personal assistant systems for health and wellness applications. His current research interests include ubiquitous sensing, pattern recognition, and embedded systems for activity and context awareness.

Martin Kusserow (S'08) received the Master's degree (with distinction) from the Eidgenössische Technische Hochschule (ETH) Zurich, Zurich, Switzerland, in 2007, and is currently working toward the Ph.D. degree in wearable computing and machine learning. For six months, he was with the Multimedia Laboratory, Toshiba Corporate R&D Center, Kawasaki, Japan, where hewas engaged in color quantization algorithms for an automotive image processing.

Mr. Kusserow received the ETH medal for his thesis on chewing sequence analysis.

Gerhard Tröster (SM'93) received the Ph.D. degree in electrical engineering from the Technical University Darmstadt, Darmstadt, Germany. He is currently a Professor and the Head of the Wearable Computing Laboratory, Eidgenössische Technische Hochschule (ETH) Zurich, Zurich, Switzerland. His current research interests include wearable computing for healthcare and production, smart textiles, sensor networks, and electronic packaging.