Recognition of dietary activity events using on-body sensors

Oliver Amft * and Gerhard Tröster

ETH Zurich, Wearable Computing Lab., c/o Electronics Laboratory, 8092 Zurich, Gloriastrasse 35, Switzerland

Summary

Objective: An imbalanced diet elevates health risks for many chronic diseases including obesity. Dietary monitoring could contribute vital information to lifestyle coaching and diet management, however current monitoring solutions are not feasible for a long-term implementation. Towards Automatic Dietary Monitoring, this work targets the continuous recognition of dietary activities using on-body sensors.

Methods: An on-body sensing approach was chosen, based on three core activities during intake: arm movements, chewing and swallowing. In three independent evaluation studies the continuous recognition of activity events was investigated and the precision-recall performance analysed. An event recognition procedure was deployed, that addresses multiple challenges of continuous activity recognition, including the dynamic adaptability for variable-length activities and flexible deployment by supporting one to many independent classes. The approach uses a sensitive activity event search followed by a selective refinement of the detection using different information fusion schemes. The method is simple and modular in design and implementation.

Results: The recognition procedure was successfully adapted to the investigated dietary activities. Four intake gesture categories from arm movements and two food groups from chewing cycle sounds were detected and identified with a recall of 80% to 90% and a precision of 50% to 64%. The detection of individual swallows resulted in 68% recall and 20% precision. Sample-accurate recognition rates were 79% for movements, 86% for chewing and 70% for swallowing.

Conclusions: Body movements and chewing sounds can be accurately identified using on-body sensors, demonstrating the feasibility of on-body dietary monitoring. Further investigations are needed to improve the swallowing spotting performance.

Key words: Automatic Dietary Monitoring, on-body sensing, activity spotting, event detection, classifier fusion, behavioural analysis, nutrition, chewing sounds
PACS:

1 Introduction

Daily dieting behaviour strongly influences the risk for developing disease conditions. The most prevalent disease associated to an imbalanced diet is obesity. Current estimations account for over one billion of overweight and 400 million obese patients worldwide. This still increasing trend was attributed to the rapid changes in society and behavioural patterns in the last decades [1]. However, obesity is not a unique diet-related disease that decreases healthy life-years in many populations. Rather, it surges the risk for related diseases, including diabetes mellitus, different types of cancer and cardio-vascular diseases. Often the diseases confound or overlay each other, preventing accurate accounting.

Several key risk factors have been identified, that are controlled by dieting behaviour. These include the timing of food intake and integration into daily schedule. For example, intermediate snacking was found to add a major part to the daily energy intake [2]. Another critical aspect is the food selection. High-energy food can be replaced by lower energy densities, such as fruits and vegetables. This improves the diet quality and lowers body weight [3].

Minimising individual risk factors is a preventive approach to systematically fight the origin of diet-related diseases. It is the most promising solution for improving quality of life in the future. Since nutrition is an inherent part of daily activities, the adoption of a healthy diet requires individual lifestyle changes. These changes need to be implemented and maintained over periods of months and years. For this purpose, a convenient long-term monitoring of dietary behaviour could become a vital tool to assess eating disorders and support diet modifications through feedback and coaching.

1.1 Dietary behaviour monitoring

No single-sensor solution exist that could capture the process of food intake and is simple to implement for diet management. Currently, dietary activities are studied manually by entering the information into food intake questionnaires. Mobile devices and Internet appliances are used to support the information entry, *e.g.* by taking pictures of the food [4] and estimating calories from entered data [5]. Further approaches to simplify data entry include the scanning of shopping receipts [6] as well as bar codes or recording voice logs [7].

These manual acquisition methods require a considerable effort of study partic-

* Corresponding author. Tel.: +41 44 632 5936; FAX: +41 44 632 1210.
Email address: amft@ife.ee.ethz.ch (Oliver Amft).

ipants, primarily to remember entering the information into the questionnaire, and study managers, to verify and analyse the data. Typically, this method is prone to errors such as imprecise timing due to back-filling, missing food item details, *e.g.* when using voice recordings [7] and low user compliance, especially for paper-based diaries [8].

Many dietary parameters such as the rate of intake (in grams/sec.) or the number of chews for a food piece are rarely assessed because adequate sensing facilities are only available in laboratory settings. However, these parameters are related to palatability, satiety and speed of eating [9]. Behavioural investigations have utilised weighting tables in controlled settings to measure the amount and rate of food intake during the consumption of individual meals [10]. An oral implant sensors was developed to acquire information about these parameters [11]. However these techniques certainly influences the user's behaviour and are not feasible for long-term monitoring.

All noninvasive dietary monitoring techniques suffer from estimation errors regarding the exact amount and calories of every consumed food item. However, a rough estimation for relevant parameters such as ratio of fluid and solid foods, food category and timing information, such as eating schedule and meal intake durations over the day, will provide a solid basis for behavioural coaching. We believe that much of this information can be extracted from on-body sensors.

1.2 Paper contributions and outline

In this work, we evaluate on-body sensing methods to automatically monitor dietary intake behaviour. In particular, three core aspects of dietary activity (*sensing domains*) were investigated by on-body sensors:

(1) Characteristic arm and trunk movements associated with the intake of foods, using inertial sensors.
(2) Chewing of foods, monitored by recording the food breakdown sound with an ear microphone.
(3) Swallowing activity, acquired by a sensor-collar containing surface Electromyography (EMG) electrodes and a stethoscope microphone.

We derive pattern models for specific activity events using the sensor data of each domain and analyse the event recognition performance. For example, individual chews are considered as events in the domain chewing. In particular, the paper makes the following contributions:

(1) We present a flexible event spotting method that can be applied either to an individual sensing modality or a combination of several. The approach

obtains its adaptivity from a variable-length feature pattern search. Its selective power originates from competitive and supportive fusion of event spottings with largely independent sources of errors. We summarise the domain-specific adaptations of the procedure. The pattern description is achieved by using time and frequency-domain features that model the temporal characteristics of an event. Using this approach, more complex algorithms, like hidden Markov models (HMMs) were avoided.

(2) We analyse the recognition of individual arm movements as well as chewing and swallowing activities from the intake of different food items. For each domain, we describe the activity sensing approach, the domain-specific recognition constraints and the conducted case studies to obtain naturalistic evaluation data. Since our work targets a combined detection and classification of the activity events, we present quantitative results for both, indicating a good performance and the feasibility of the sensing approaches for Automatic Dietary Monitoring.

The evaluations are performed on data from three different studies. To analyse the recognition performance under realistic conditions, the data sets included other common activities, *e.g.* conversations and arbitrary movements.

2 Dietary activity domains and related work

Activity monitoring and recognition has attracted researchers from many backgrounds, including machine vision and more recently pervasive and wearable computing. An exhaustive review of the literature is beyond the scope of this work. Instead, we focus on systems for behaviour and Automatic Dietary Monitoring as well as research on the three sensing domains considered in this work.

Approaches towards Automatic Dietary Monitoring typically build on intelligent infrastructures. Chang et al. [12] developed a monitoring table to detect activities in a dining scenario. The table is partitioned into several sensing sections equipped with radio-frequency-identification (RFID) readers to identify food containers and weight sensors to track food transport between containers and personal plates. The precision of the system is bound to the spatial resolution of table sensing sections and requires static assignment of food containers to these sections. The concept of load sensing on a table surface for user activity detection was introduced earlier by Schmidt et al. [13]. In their approach coarse object movements were estimated from a single sensing section.

Beigl et al. [14] equipped household objects with sensing capabilities. In the presented example, a cup was chosen to identify activities carried out with it.

For dietary monitoring applications, RFID technology has great potential as a combined wearable and environmental sensing modality. Patterson et al. [15] attached tags to 60 household objects. The detection was restricted to morning activities, recorded by an RFID reader worn at the user's hand. The activities included, using the bathroom, preparing breakfast foods and eating breakfast.

The infrastructure sensing approaches provide valuable information on various user activities were sensors can be easily attached or hidden. However the approaches generally suffer from the user identification problem: while one user may prepare the foods, several others can consume them. Wearable sensors can bridge this gap and associate the user directly to the activities. Moreover, since worn at the body, the sensors can reveal more detailed information that otherwise would require laboratory setups.

2.1 Movement recognition

Movements and gestures related to dietary intake can be roughly discriminated into a preparation phase of the food or beverage items, such as unpacking, opening, cooking and plate or cup filling, and the actual feeding. The feeding movements target the fine-cutting, loading, and manoeuvring of the prepared piece to the mouth. In the feeding phase specific tools, such as fork and knife can be used.

Our focus is to recognise intentional arm and upper body movements during the feeding phase. These movements are a result of handling the tool in the hand(s) and the food material properties viscosity and size. These properties relate directly to the food category. For example a soup is usually feed with a spoon while a glass, cup, or bottle is used for drinking. Hence all relevant movement events can be characterised as directed gestures of the left or right arm, supported by the upper body.

A large base of existing works addressed the problem of classification on well-defined sequences or previously isolated gestures, *e.g.* for Kung Fu moves [16] or in a worker assembly scenario [17]. Works that targeted the continuous recognition used explicit segmentation steps or implicit segmentation capabilities of algorithms, such as HMMs. Lee and Kim [18] used HMMs and introduced a threshold model to eliminate detection noise. The threshold model is constructed from all trained gesture models. Explicit segmentation was used by Ward et al. [19] in an assembly task. Recognition was achieved by fusing classifier outputs. Lee and Yangsheng [20] used acceleration thresholds in combination with HMMs. In previous works of the authors on intake gesture recognition, HMMs were used together with an explicit data-adaptive segmentation [21].

While HMMs are helpful to model the temporal structure of movements, they were avoided in this work to minimise the complexity of the search procedure for both training and actual search.

2.2 Chewing recognition

Chewing targets simultaneous food breakdown and lubrication to form a food bolus that can be swallowed. A chewing sequence starts after the food piece is transferred to the mouth. The food breakdown is composed of arbitrary tongue movements and cyclic opening and closing of the jaw (*chewing cycle*). During the material breakdown sounds are emitted that are partially audible by air-conduction in the near vicinity, but effectively transmitted by bone-conduction from teeth and jawbone to the skull and the ear canal.

The emitted sounds are related to the food material texture. Interaction of chewing with the acoustic sensation and perception of food items has been investigated to study food preferences. Typically, studio recording setups were used to analyse air-conducted chewing sounds [22] and laboratory installations to assess the deformation sounds with a destruction instrument [23]. The loudness of a food item during chewing depends mainly on its inner structure, the arrangement of cells, impurities and existing cracks [24]. Wet cellular materials, such as apples and lettuce, are termed *wet-crisp* since the cell structures contain fluids, whereas *dry-crisp* products, such as potato chips have air inclusions [25].

The food deformation in a chewing cycle is understood as a gradually decomposition of the material structure, observed as a decline of the sound level [26]. Initial attempts were made by DeBelie et al. [27] to discriminate two classes of crispness in apples by analysing principal components in the sound spectrum of the initial bite. In a followup work DeBelie et al. [28] classified the sound emissions from the initial bite of different dry-crisp snacks. Both works addressed the isolated classification. In our previous work the microphone positioning and classification of four different foods was investigated [29]. The ear canal provided the best signal (chewing) to noise (user speaking) ratio. This sensor positioning can be comfortable and socially acceptable for continuous monitoring, comparable to mobile headsets or hearing aids.

In this work, following our recognition approach, the identification of individual chewing cycles from food breaking sounds was targeted. The food category is subsequently classified from the sound pattern of the cycle.

2.3 Swallowing recognition

Swallowing is a frequent activity during food intake. It is mostly performed unconsciously and when initiated, controlled by a pattern of muscle activations [30]. The swallowing act is often partitioned into (1) oral preparation phase (food in the mouth), (2) pharyngeal phase (food bolus in the throat) and (3) esophageal phase (food propulsion towards the stomach) [31]. After transforming the food to a allowable bolus in the oral phase, the swallowing reflex is initiated by the tongue, starting the pharyngeal phase. In this phase a sequence of muscle activations is used to transport the bolus and protect the respiratory tract.

A number of clinical assessment methods have been developed to analyse the complex interaction of swallowing, phonation and respiration at the pharynx and diagnose abnormal swallowing in the pharyngeal phase. The assessment methods can be broadly grouped as invasive methods, that require a strict laboratory or clinic setting and a variety of non-invasive sensing methods. In the latter category, the following main approaches were taken: sensing muscle activations by surface EMG, *e.g.* [32], listening to the throat sounds using a stethoscope [33] as well as stethoscope-like acoustic transducers or sealed microphones [34].

A large share of research works targeted the basic understanding of the swallowing process, only few addressed the continuous monitoring. Danbolt et al. [35] used sensors to detect hyoid movement at the throat. It was found that the sensor incurs heavy measurement artifacts from neck and tongue movements as well as from speaking. Limdi et al. [36] tracked muscle contraction intensity based on surface EMG to inform the user of elevated swallowing rates. Sukthankar and Reddy et al. [37] used surface EMG and vibration sensors and targeted applications in dysphagia rehabilitation. Both latter works did not present a performance evaluation for their approaches to the continuous recognition problem. In our previous work [38], swallowing was analysed from surface EMG and sound for the isolated classification of swallowed bolus types, *e.g.* solid or fluid. Moreover, an initial investigation towards the continuous detection was made. The approach is taken forward in the present evaluation by extending the swallowing study and evaluating the performance of different fusion methods.

3 Recognition and evaluation methods

The envisioned system shall be continuously worn during daily routine. In all sensing domains relevant activity events occur only sporadically, often embed-

ded into a large set of other, non-relevant activities (*NULL class*). For example, stethoscope-like sound recordings intended to record swallowing sounds at the throat, inherently pick up speaking, or even environmental noises.

A method that targets the spotting of relevant activity events should be effective in retrieving correct events while omitting NULL class data. However, the sensing domains considered in this work have very few constraints, resulting in a highly variable NULL class. As a consequence of this diversity, it is not feasible to derive a model for NULL (garbage model) without integrating assumptions about these random activities. Moreover, training of the relevant event model(s) should be critically reviewed for its dependency on NULL.

Another challenge is the variable length of the activities, leading to duration variances in the relevant events. Consider for example a intake gesture using fork and knife where the food must be cut into appropriate sized pieces before manoeuvring it to the mouth. This indicates that a simple, fixed sliding window search would not be able to identify the gestures accurately.

Our approach to detecting and classifying dietary activities is based on three main steps: (1) an explicit segmentation of signals to define search bounds, (2) a sensitive event detection using a feature similarity search algorithm with an adaptive, dynamically defined window size, and (3) a selective fusion of detection results exploiting independent sources of error to filter out false positives and obtain an event classification in the same step. Figure 1 outlines the components of our event detection and classification method.

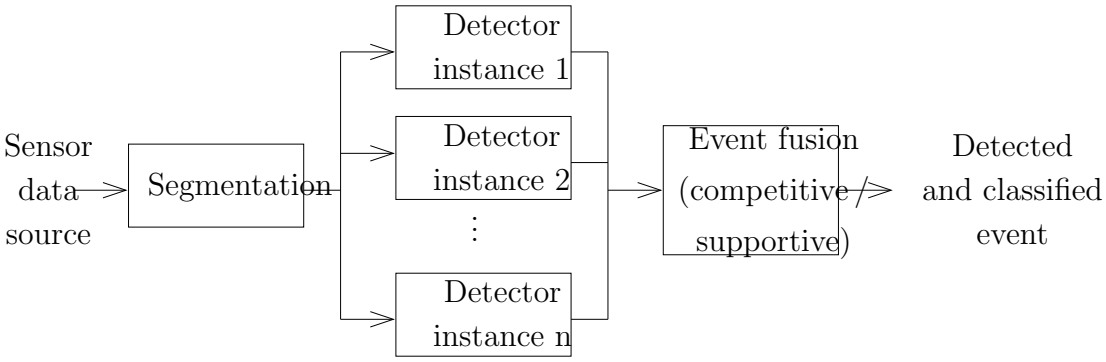

Fig. 1. Event detection and classification procedure used in the work. The detector instances (1 to n) can be trained to spot activity event patterns of specific classes or individual modalities. The event fusion can combine events of different type (competitive) or modalities for one type (supportive). Both concepts are presented in this work.

3.1 Event recognition procedure

In the first step a segmentation is obtained that specifies the bounds for the following search. Various data-adaptive methods or a fixed distance can be used for this purpose. In this work, we used the latter approach with a domain-specific distance setting.

3.1.1 Event detection using feature similarity search

The event detection step utilises the segmentation points to search for potential activity event sections using a similarity-based algorithm. The search is performed by comparing features of a data section under investigation to a previously trained pattern.

The following search principle is illustrated in Figure 2. For a given segmentation point, the history of sensor data is analysed between a lower and upper search bound. These bounds are determined in the training step from the overlapping of manually annotated events and the segmentation points. For each search section the similarity of a feature set to a pre-trained set is quantified by computing the Euclidean distance (D_{Event}) between them. A distance threshold (D_{Thres}), also obtained during the training, is used to remove unlikely sections. The similarity search works as a detector that returns a list of event sections associated with a distance to the training pattern.

One benefit of this algorithm is that it can operate as a single pattern detector, when applied to retrieve one relevant type from continuous sensor data only. Using the feature similarity search, multiple detector instances can be combined to independently spot different classes. This permits an independent feature set for each class. Furthermore, as we will show for the detection of swallowing, instances trained from independent sensing modalities can be used to detect one event type in parallel.

3.1.2 Competitive and supportive event fusion

By selecting an appropriate distance threshold (D_{Thres}), the similarity search is configured to spot most of the activities in the sensor data. Consequently it can incur false positives. In the fusion step different class- or modality-specific event detectors are combined to reduce these errors. This improvement originates from the independent sources of error of each detector and modality.

For multiple detectors a competitive fusion strategy was used to select the final events. A supportive strategy was deployed to combine the modality-specific detection of one activity type, since here the detectors could reinforce each

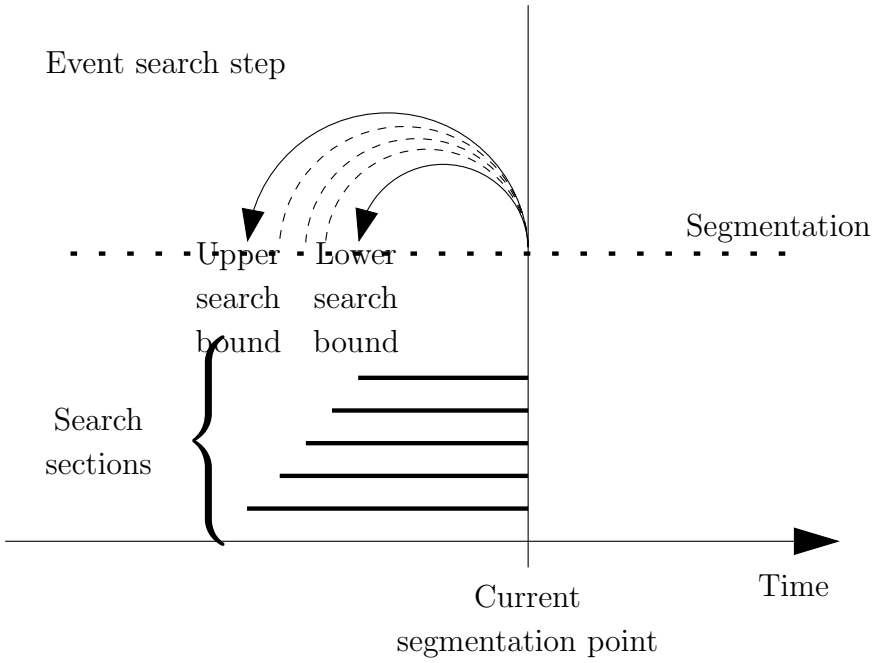

Fig. 2. Schematic of the activity event search step. The segmentation is indicated by the dotted line. The search is performed by computing feature sets from the sensor data (not shown) between lower and upper search bounds. The search sections are evaluated by comparing their feature sets to a pre-trained pattern. (Please refer to the text for more details.)

other.

In this work we evaluated different fusion methods: (1) comparison of the events, keeping the event with the highest confidence (COMP), (2) agreement of the detectors (AGREE) and (3) re-weighting of the detection by logistic regression (LR). The methods are commonly used to combine classifier outputs [39,19]. In this work, COMP corresponds to the competition strategy and AGREE implements a supportive approach. LR can be used for both strategies.

To select the most probable from concurrently reported events, the competitive fusion compares a confidence associated to each event. This confidence was derived from the similarity search distances (D_{Event}) by normalisation using the distance threshold (D_{Thres}) in each detector instance (Equation 1).

$$Confidence = \frac{D_{Thres} - D_{Event}}{D_{Thres}} \tag{1}$$

A sliding buffer of candidate events is used and continuously updated as new events are entering from the detector instances. For each entering event the collision (temporal overlapping of the event section with events already in the buffer) is resolved according to the selected fusion strategy. The events are

released from the buffer after a timeout as final result of the procedure.

3.2 Feature computation

The temporal structure of many complex activities is a key element for their pattern modelling and subsequent machine recognition. For example, movements are frequently modelled with HMMs and time-continuous features to capture this effect.

In this work, we integrated the temporal structure of the activity events in individual single-value features. The features were computed for predefined sections of an event. We spitted the event in two or four slices. This solution provided an acceptable trade-off between temporal description and total number of features. The solution permits a combination of sliced features and features for the entire event. Moreover, this approach can simplify both modelling and event search, compared to time-continuous features. We used it with the recognition approach presented above. The similarity search is then performed using the features to describe each event and search every section.

3.3 Evaluation procedure

3.3.1 Experimental concept

The analysis of each sensing domain was based on experimental data, individually acquired for each domain. Figure 3 indicates the sensor attachment at the body for all domains. For the recording of movements a commercial motion acquisition system based on inertial sensors was used. Customised systems were utilised for the chewing (ear microphone) and swallowing (sensor collar) recordings. Table 1 provides a detailed description of the sensors used. In each study the activities were manually annotated by an observer. The study procedures are further detailed in the evaluation sections for each sensing domain.

3.3.2 Soft alignment procedure

In order to account an event as recognised, the detection procedure must return a valid begin and end of an activity section and its identity (for multi-class detections). The section boundaries were compared to begin and end of the annotated events. However the boundaries do not match exactly since the manual annotation was not accurate on the granularity of each sample

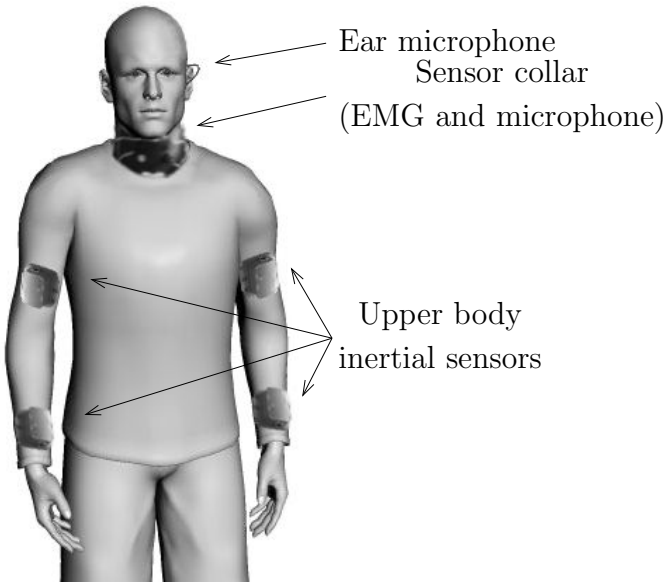

Fig. 3. Schematic sensor positioning at the body. (See Table 1 for a detailed description.)

and the segmentation algorithm can introduce a small alignment error in the detection.

For the feasibility in the envisioned dietary monitoring application the exact alignment is not a critical aspect, if the event is associated to the true activity at all. Hence, we applied a soft alignment matching, following the concept of a boundary jitter. Equation 2 describes the accounting of correct events.

$$Recognised = \begin{cases} \text{true, if } j \leq max\left(\frac{|A_{Begin}-E_{Begin}|}{A_{End}-A_{Begin}}, \frac{|A_{End}-E_{End}|}{A_{End}-A_{Begin}}\right) \\ \text{false, otherwise} \end{cases} \quad (2)$$

The parameters A_{Begin} and A_{End} correspond to start and stop sample of the manual annotation and likewise, E_{Begin} and E_{End} to the retrieved event. The jitter parameter j can be set, depending on the acceptable jitter for an application. The jitter $j = 0$ corresponds to an exact matching of the boundaries and $j = 1$ would allow a jitter in size of the event duration. Moreover, this accounting procedure assures that large events, covering more than the annotation section, will be rejected as well, if their begin and end do not conform to Equation 2. Multiple counts of matches and misses were especially avoided.

For the evaluation in this work a jitter of $j = 0.5$ was chosen. We believe that this is an adequate accuracy for applications in dietary monitoring.

Table 1
List of sensors systems used in the dietary activity studies.

Sensor type	**Sensor description**	**Sensing domain**
Inertial sensors	Sensor modules containing acceleration sensors, gyroscopes (rate of turn) and compass sensors (magnetic field), each in 3 dimensions. The modules were attached to the user's arms. Manufacturer: XSens, model: MTi.	Movement activity
Ear microphone	Electret miniature condenser microphone. The microphone was embedded into an ear pad foam and worn at the ear canal. Manufacturer: Knowles Acoustics, model: TM-24546.	Chewing activity
Stethoscope microphone	Electret condenser microphone. The microphone was attached with medical tape or worn in a collar below the hyoid. Manufacturer: Sony, model: ECM-C115.	Swallowing activity
Electromyogram (EMG)	Electromyogram electrodes and acquisition system. Electrodes were directly attached or worn in a collar at the infra-hyoid throat position. Manufacturer: MindMedia, model: Nexus-10.	

3.3.3 Performance measurement

To account for variations in the acquired data sets, a four-fold cross-validation procedure was used to determine training and testing set for the performance analysis. For training, three of four data parts were used. Evaluation was performed on the left-out data part. This procedure was repeated until all four parts were used for testing once. The partition boundaries were adapted to avoid intersecting the manually annotated event sections. The choice of four partitions reflects an empirical trade-off between processing effort, the need for enough training observations in all combinations of the partitions and the intended averaging effect for the final results. An additional performance gain could be achieved by higher iteration counts, potentially using more events for training.

To analyse the recognition performance, we used the metrics *Precision* and *Recall*, commonly used for information retrieval assessments. These metrics are derived as follows:

$$Recall = \frac{\text{Recognised events}}{\text{Relevant events}}, \qquad Precision = \frac{\text{Recognised events}}{\text{Retrieved events}} \tag{3}$$

Relevant events corresponds to the manually annotated number of actually occurred event instances. *Retrieved events* represents the number of events returned by the event recognition procedure. Finally, *Recognised events* refers to the correctly returned number of events. Both metrics have a value range of $[0, 1]$. A recall value of one indicates a perfect accuracy of a method (all relevant events are recognised), while a precision value of one indicates that the method does not return false positives (insertion errors).

4 Movement recognition

4.1 Study description

To evaluate our recognition approach for movements, a case series was recorded, utilising commercially available inertial sensors. Table 1 specifies the sensors used. The inertial sensors were attached onto a jacket at the lower and upper arm as well as the upper back. Figure 3 illustrates the sensor positions.

The movements of the arms and upper body was recorded with a sampling rate of 100 Hz from four right-handed volunteers (1 female, 3 male, aged between 25 to 35 years). The participants were seated in front of a table carrying the food items and tools. They were instructed to eat and drink as they would normally do.

Intake sessions were recorded from each participant on separate days. Four intake activities were recorded for each session: (1) eating meat lasagne with fork and knife (cutlery, CL), (2) fetching a glass and drinking from it (DK), (3) eating a soup with a spoon (SP), and (4) eating slices of bread with one hand only (HD). All meals were served at adequate temperature for normal eating/drinking. Table 2 summarises the acquired data which was inspected and annotated.

In order to enrich diversity of the data set and avoid long periods without movements, the participants were asked to conduct a set of other, non-relevant movements and gestures. Besides arbitrary movements of the participants the following additional arm gestures have been recorded and annotated to quantify the data set noise: scratching head (96 times), touching chin (92 times), reading and turning pages of newspaper (99 times), using tissue (89 times), glancing at the watch (92 times) and answering a simulated mobile phone call (90 times), all total numbers of the data set.

Table 2
Movement study: Statistics of acquired and annotated intake gestures.

Number of participants	4
Annotated gestures	1020
Relevant event share	97.44 min (34.7%)
Total length of data set	4.68 hours

4.2 Evaluation results

The event recognition procedure was adapted to the movement domain in the following way:

(1) A time constant of 0.5 s was used for segmentation.
(2) For each of the four gesture categories an event detector instance was trained. Using the Euler angles of the lower arms, features such as mean, variance and signal sum in four sliced sections and for the complete gesture were computed. By visually inspecting test recordings we found that the upper arm and the back sensors could not support the recognition without constructing a more complex body model. Hence, they were excluded from the analysis.
(3) The event fusion using the competitive strategy was subsequently applied to the detector instance results and the event category with the highest confidence was selected as final result. Due to variable lengths of gestures in our data set, the candidate buffer was configured to release events only after 30 s.

Figure 4 shows precision-recall (PR) graphs for a user-specific evaluation of the movement event fusion using the COMP method. The curves were created by evaluating the performance at various confidence thresholds for every class and for every participant (A-D). Best performance is found towards the top-right corner (high precision, high recall).

Both graphs indicate a good performance for the movement event recognition. The best result was achieved for the category DK, while HD performed less well. Since the latter gesture is very simple it was often confused with other movements towards the head. In contrast, DK is more complex (fetching, drinking). The second graph shows that all participants performed similarly well.

Table 3 summarises the results obtained from the event detection and the event fusion. For the SP gestures, we observed that participants bend themselves over the bowl, to avoid spilling and to minimise the movements. This affected the detection performance, since only lower arm features were used in the

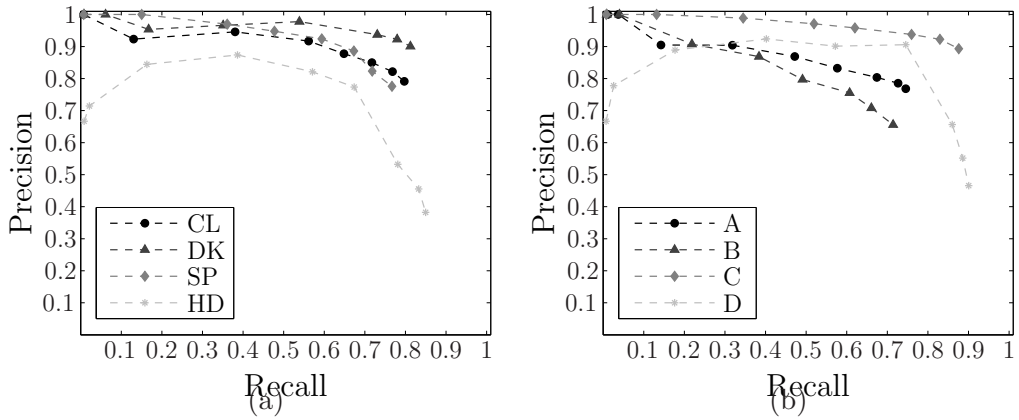

Fig. 4. Movement study: User-specific PR analysis (confidence threshold sweep) of the event fusion results using the COMP method. Best performance is found towards the top-right corner (high precision, high recall). (a) Analysis for every category (CL=cutlery, DK=drink, SP=spoon, HD=hand only). (b) Analysis for every study participant (A-D).

evaluation.

Table 4 shows a confusion matrix of the event recognition, obtained by comparing the recognition results to the annotation for each sensor data sample. Complementary to the soft alignment counting scheme used for the results in Table 3, this representation shows the sample-accurate result. For all categories and NULL a recognition rate of 75% to 82% was achieved. This rate was computed as class-relative accuracy ($\frac{correct_C}{relevant_C}$).

Table 3
Movement study: Summary for the user-specific performance for the event detection and the fusion method COMP.

Metric	**Event detection**				**Event fusion (COMP)**				
	CL	DK	SP	HD	CL	DK	SP	HD	**Total**
relevant	276	245	266	233	276	245	266	233	**1020**
retrieved	347	247	284	717	278	221	263	518	**1280**
recognised	223	210	208	201	220	199	204	198	**821**
deletions	53	35	58	32	56	46	62	35	**199**
insertions	124	37	76	516	58	22	59	320	**459**
recall	0.81	0.86	0.78	0.86	0.80	0.81	0.77	0.85	**0.80**
precision	0.64	0.85	0.73	0.28	0.79	0.90	0.78	0.38	**0.64**

Table 4
Movement study: Confusion matrix of the final user-specific evaluation result using COMP fusion (duration in seconds and ratios).

		Predicted category				
		NULL	CL	DK	SP	HD
Actual category	NULL	8869	613	233	305	982
		(81%)	(6%)	(2%)	(3%)	(9%)
	CL	452	2130	0	0	8
		(17%)	(82%)	(0%)	(0%)	(0%)
	DK	302	1	1182	0	34
		(20%)	(0%)	(78%)	(0%)	(2%)
	SP	237	19	0	807	10
		(22%)	(2%)	(0%)	(75%)	(1%)
	HD	103	20	0	0	541
		(16%)	(3%)	(0%)	(0%)	(81%)

5 Chewing recognition

5.1 Study description

For the evaluation of chewing sounds we used an ear microphone as indicated in Figure 3. The miniature microphone was build into a standard type ear pad and kept at the ear canal by an ear hook, as it is used for mobile phone headsets. In a single case study the chewing sounds from different foods were recorded at 16 bit, 44 kHz from a male individual with natural dentition (aged 29 years).

The participant was seated conveniently on a chair close to a table carrying the foods. He could still hear normal-level conversation in the room and was allowed to move and speak during the recording sessions. The room was controlled for a constant noise level of an office environment (the recording in a sound studio was avoided). Recordings were made in individual sessions on separate days. The participant took bites from the foods as he wished. All of the foods belonged to his normal diet. The food products included for the recognition analysis were:

(1) Dry-crisp food: potato chips, approx. 3 cm in diameter

(2) Wet-crisp foods: (1) mixed lettuce, containing endive, sugar loaf, frisée, raddichio, chicory, arugula, and (2) raw carrots.
(3) Soft foods: (1) cooked chicken meat and (2) pasta.

The foods evaluated in this work, contained many chewing cycles. Manual annotation of every chewing cycle was performed in a post-recording step by reviewing the waveforms and listening to the sounds. This procedure is accurate in identifying every chewing cycle until the food bolus is swallowed, however it makes the recordings very expensive.

The recordings included chewing sounds from further food products (bread and chocolate), as well as environmental conversation and speaking. Table 2 summarises the acquired data which was inspected and annotated.

Table 5
Chewing study: Statistics of acquired and annotated chewing sounds.

Number of participants	1
Annotated chewing cycles	1947
Relevant event share	10.50 min (21.7%)
Total length of data set	0.81 hours

5.2 Evaluation results

The event recognition procedure was adapted to the chewing domain in the following way:

(1) A time constant of 125 ms was used for segmentation. This choice was made based on the average duration of a chewing sound (as annotated) of 350 ms or less, depending on the food type.
(2) Initially, for each of the three food categories a feature similarity instance was trained. Using the microphone data, spectral features such as band energy, auto-correlation and cepstral coefficients in four sliced sections were computed. We observed during the evaluation, that the detector for soft foods worked poorly, resulting in many insertion errors. This behaviour was attributed to the low signal to noise ratio. We omitted this model in the further evaluation to demonstrate the good performance of the dry and wet food detectors.
(3) The event fusion using the competitive strategy was subsequently applied to the detected chewing cycles and the category with the highest confidence was selected as final result. We analysed the COMP and LR methods for the fusion.

The low-amplitude chewing sounds from the soft foods (meat and pasta) created a special problem for the detector. While a high recall was achieved, the detection was very sensitive to other sounds (as seen in the low precision in Table 6). COMP and LR fusion of the three detectors did not solve this problem, because the number of soft-food insertions was too high.

For every intake cycle all chews were annotated until the food bolus was swallowed and the normal mouth cleaning phase began. In this phase, chews were hard to observe in the sound waveform. However the algorithm was still able to detect them. Figure 5 visualises an example waveform including a chewing sequence of potato chips, the cleanup and a conversation phase. For this food the chewing cycles can be seen very well in the sound waveform. The vertical bars indicate the annotation. In the lower plot, the detected chewing events are shown as horizontal bars. As the diagrams shows, additional events were reported for the cleanup phase. We exemplarily verified that these chews were correctly retrieved.

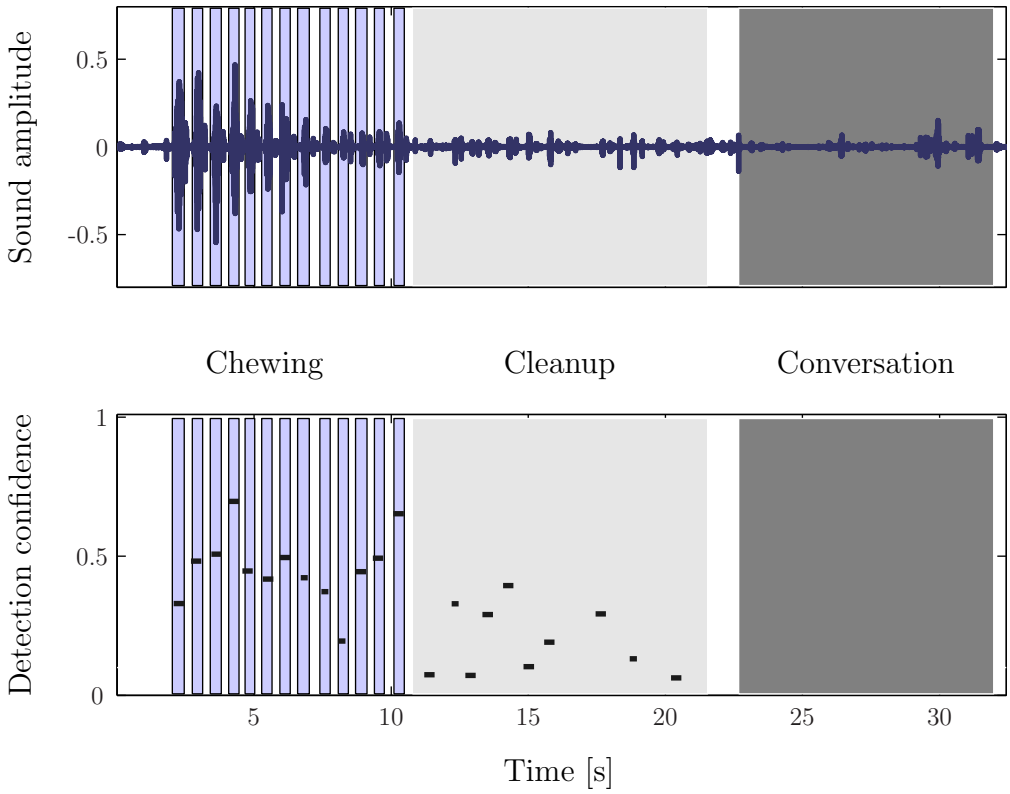

Fig. 5. Chewing study: Example waveform of a chewing sequence of potato chips, cleanup and conversation phases, indicated by the shaded areas. Upper plot: sound waveform. Lower plot: chewing cycle detection result. (The detector correctly identified chewing cycles in the cleanup phase, that were not annotated. Please see the related text for more details.)

Since the actually existing chews in the cleanup phase could not be automatically verified, they were counted as insertion errors. The impact can be seen

in the PR performance analysis in Figure 6 and the summary in Table 6. For both food categories the COMP and LR fusion methods return good results. We concluded from the quantitative summary in Table 6 that LR removes slightly more insertion errors and has less deletions.

Table 7 shows the confusion matrix derived by applying the LR method. Using the same procedure as presented for the movement confusion analysis, class-relative recognition rates of 85% to 87% were achieved. This indicates a very good performance. Especially, a low confusion rate of the dry and wet categories was observed.

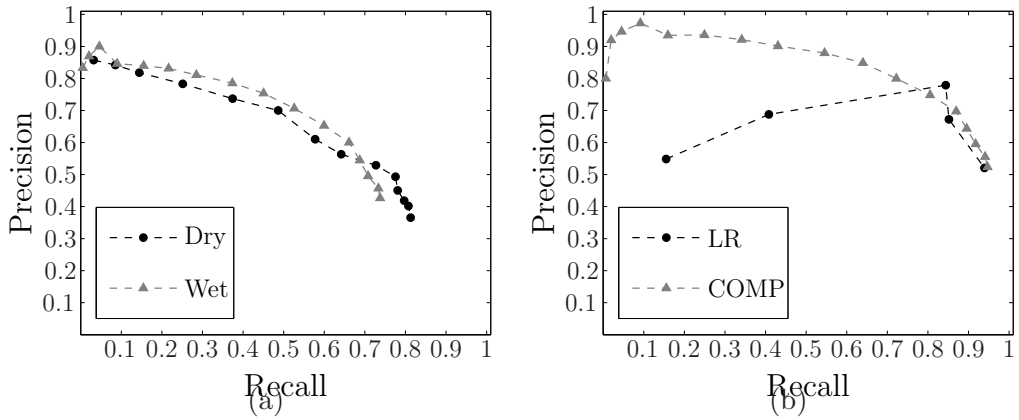

Fig. 6. Chewing study: User-specific PR analysis (confidence threshold sweep) of the event fusion stage. Best performance is found towards the top-right corner (high precision, high recall). (a) Analysis for the two food categories ("dry" and "wet"). (b) Analysis for the two competitive fusion methods (COMP and LR).

6 Swallowing recognition

6.1 Study description

Swallowing was analysed from surface EMG electrodes and a microphone sensor. The sensor positioning was equal for all participants. For some participants the sensors were embedded in a collar. The collar helped to quickly attach the sensors to the correct throat region. The location of the EMG was constantly verified, however the collar supported the stable positioning at the infra-hyoid position very well. The microphone was situated at the lower part of the throat, below the larynx. EMG was recorded at 24 bit, 2 kHz and band-pass filtered. Sound data was recorded at 16 bit, 22 kHz. Figure 3 and Table 1 summarise positioning and setup of the sensors and the collar.

Six volunteers (4 male, 2 female, aged 20 to 30 years) without known swallow-

Table 6
Chewing study: Summary for the user-specific performance for the event recognition (three categories) and the fusion methods (COMP and LR). The fusion results were derived using the food categories "Dry" and "Wet" only.

Metric	**Event detection**			**Event fusion**					
				COMP			**LR**		
	Dry	Wet	Soft	Dry	Wet	**Total**	Dry	Wet	**Total**
relevant	187	979	781	187	979	**1166**	187	979	**1166**
retrieved	1327	2098	3483	416	1693	**2109**	416	1687	**2103**
recognised	186	909	460	152	722	**874**	184	900	**1084**
deletions	1	70	321	35	257	**292**	3	79	**82**
insertions	1141	1189	3023	264	971	**1235**	232	787	**1019**
recall	0.99	0.93	0.59	0.81	0.74	**0.75**	0.98	0.92	**0.93**
precision	0.14	0.43	0.13	0.37	0.43	**0.41**	0.44	0.53	**0.52**

Table 7
Chewing study: Confusion matrix of the final user-specific evaluation result using LR fusion (duration in seconds and ratios).

		Predicted category		
		NULL	Dry	Wet
Actual category	NULL	2791 (86%)	100 (3%)	344 (11%)
	Dry	12 (13%)	76 (87%)	0 (0%)
	Wet	57 (15%)	3 (1%)	332 (85%)

ing abnormalities were instructed to eat and drink different food items: 5 and 15 ml of water, spoonfuls of yoghurt and pieces of bread (approx. 2 cm^3). The individuals were seated conveniently on a chair in front of a table carrying the foods. They were allowed to move, chew and speak normally during the recording sessions. The room was controlled for a normal and constant noise level of an office environment. To account for physiologic variations, two intake sessions were recorded on different days. The participants were asked to swallow the food items in one piece after chewing and manipulating the bolus as usual. None of the participants expressed a dislike for any of the included foods nor problems to swallow the selected bolus sizes. Table 8 summarises

the acquired data that was inspected and annotated.

Table 8
Swallowing study: Statistics of acquired and annotated swallowing activity.

Number of participants	6
Annotated swallows	1265
Relevant event share	44.58 min (9.3%)
Total length of data set	7.93 hours

6.2 Evaluation results

The event recognition procedure was adapted to the chewing domain in the following way:

(1) A time constant of 250 ms was used for segmentation.
(2) Feature similarity instances were trained using the EMG and microphone data individually. The foods were initially grouped regarding their expected bolus size into small (5 ml water, spoonfuls of yoghurt and pieces of bread) and large (15 ml water). This approach was dropped, since no clear discrimination of the two categories was found. In the following, we targeted the detection without further classification. We concluded from early tests that the EMG is disturbed by different muscle activations, independent from swallowing. The investigated hyoid muscle is covered by several layers of other muscle tissue. We concentrated on a simple activity detection using time domain features such as sum, maximum and peaks of the signal. For the sound data, spectral features such as band energy, auto-correlation coefficients and signal energy were used. An initial test of sliced features did not lead to an improvement in recognition.
(3) The event fusion using a supportive strategy was subsequently applied to the detected swallowing events from EMG and sound data. We analysed the performance of AGREE and LR methods.

For the AGREE fusion all participants reached a high recall, indicating that the detection procedure was able to retrieve many events. Figure 7 presents the corresponding PR analysis. The evaluation revealed two groups: for participants (C and D) the detection performance was higher than for the others. However, these participants did neither belong to the same gender, nor were they recorded with the collar. We observed that many other participants exhibited either a high EMG response or sound, for C and D both sensors provided a consistent event pattern. Consequently, both EMG and sound-based detection more often returned a correct result for them, whereas for the remaining participants no reduction of the insertion errors was achieved. Further investigation of this issue is required.

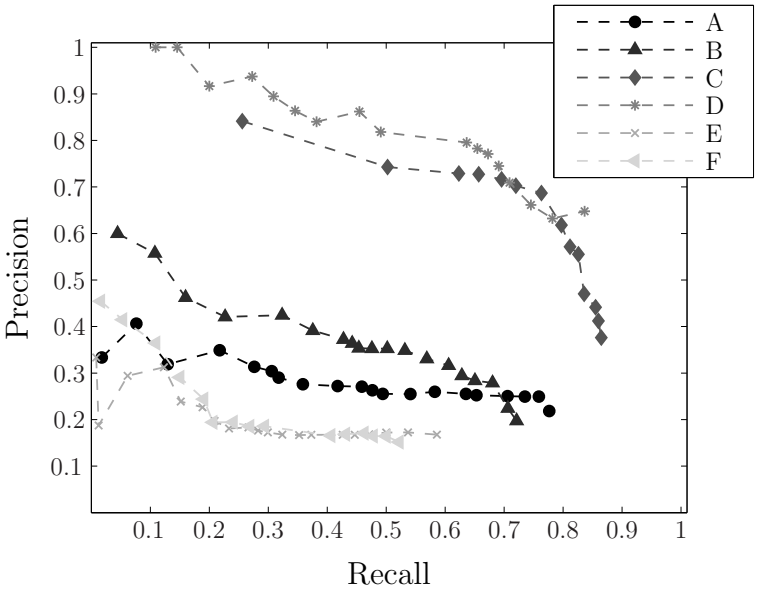

Fig. 7. Swallowing study: PR analysis (confidence threshold sweep) for each study participant (A-F) using the agreement fusion (AGREE). Best performance is found towards the top-right corner (high precision, high recall).

On average for all participants, the AGREE fusion method improved the precision. LR did not improve the individual spotting results. Table 9 summarises the results obtained from the event detection instances and the fusion methods.

The sample-accurate detection result was determined from the AGREE fusion result. The swallowing recognition rate was 64%, for the NULL class 75% were obtained. This indicates that the detection provides a sensible result.

7 Discussion

7.1 Methodology

The continuous recognition of dietary activity events from sensor data patterns was evaluated in this work. Spotting activity events in continuous sensor data is a vital prerequisite for the deployment of activity detection in general. While the targeted activities can be described by a domain expert, the embedding data (NULL class) cannot be modelled due to the degrees of freedom in the human activities and the cost for large training data sets. Consequently, assumptions about the embedding should be minimised to achieve an acceptable performance generalisation. We believe that the current work is a step towards resolving this challenge, although the presented method is not completely free from assumptions. The most critical aspects in this respect include

Table 9
Swallowing study: Summary for the user-specific performance for the event detection using muscle activity (EMG), audio (SND), and the fusion methods (LR and AGREE).

Metric	**Event detection**		**Event fusion**	
			LR	**AGREE**
	EMG	SND	EMG+SND	EMG+SND
relevant	1265	1265	1265	1265
retrieved	6046	8093	8085	4345
recognised	955	834	824	861
deletions	310	431	441	404
insertions	5091	7259	7261	3484
recall	0.75	0.66	0.65	0.68
precision	0.16	0.10	0.10	0.20

the selection of features and event detection thresholds.

A combination of individual single-value features for activity event slices were used for the detection. With this approach the temporal structure of the activities was transformed into a spatial representation. This is a useful concept to model activities for the continuous search. In an earlier work, we applied this principle to the recognition of gaming gestures only [40]. For each domain, features were selected from visual inspection of the sensor waveforms and from previous experience. We expect that the recognition performance could be improved by a thorough feature search and selection strategy. This will also help to identify sensors that can be omitted or adjusted in its placement.

We introduced the scheme of competitive and supportive event fusion to construct a selective refinement step for spotted events. By design of the recognition system, the choice of the fusion strategy is made. The supportive strategy was applied for spottings from independent sensors, describing the *same event type*. Using competitive fusion, we selected the most appropriate event from *different event type* spottings. Both strategies could be combined to more complex selection schemes. In related works, they have been used to combine classifier outputs mostly [19].

An advantage of our method is its ability to work on single event detection classes with individual feature sets. For the detection of one event type, typically a supportive fusion strategy can still be used, by deploying different sensors. An application for detecting single event types in dietary monitoring was shown in the swallowing evaluation. Further applications are the detection

of drinking gestures to assess fluid consumption or using a single food model to assess one category of foods in dietary intake.

In order to describe the complexity of the event detection as a search problem, we listed the embedding size of the data sets. This size was expressed as ratio of total annotated event duration over the total length of the data set. For the data sets in this work, the ratio was 34.7% for the movement, 21.7% for chewing and 9.3% for the swallowing study. The ratio indicates the severity of the search: the smaller the ratio, the more difficult it is to achieve a good recognition results due to the large and potentially diverse embedding data. However, we believe that the high embedding size in the swallowing study is not the unique reason for its weak precision. Section 7.4 discusses the swallowing study in detail.

We introduced a soft alignment measure to account for the variability in alignment between annotation and event detection. A boundary jitter normalised by the annotated length of the event was defined as threshold, below which the event is counted as recognised. The larger the jitter, the more mismatch in alignment is allowed and an event reporting that may otherwise be accounted as insertion/deletion will be accepted as correct. In its extreme, the counting of correct events could be made by simply checking if an overlap with the annotation exist at all. For the targeted applications in dietary monitoring an exact match is less critical as long as the activity is captured at all. Therefore, we selected a jitter value that is neither too optimistic (by permitting large alignment errors) nor pessimistic (being overly strict in the boundary match). The comparison with sample-accurate confusion matrices confirms that the soft alignment is a sensible solution for event spotting performance analyses. For a more detailed analysis of detection errors, the Error Distribution Diagrams [41] could be used.

7.2 Movement recognition

Different gesture types were defined, that occur frequently in European and American diets, to evaluate the recognition of food intake movements. The results indicate that all types could be recognised from lower arm motion, most of them with good accuracy. To improve the recognition of certain gestures, information from inertial sensors at the subject's back could be added. The proposed event fusion method is a valuable addition to the feature similarity search for movement detection. In a related work of the authors, a two-stage approach based on a similarity search and HMMs was used [21]. While the HMMs proved valuable for refining the detection result in the second stage, they add a high complexity in both, initial design and parameter estimation. In comparison, the performance achieved with the event fusion approach in

the current work could match the recall, but performs approx. 10% lower in precision than the HMMs on the same data set. Further refinement of features and segmentation could close this gap. Moreover, we presented a rigorous evaluation framework using cross-validation in this work, that was not previously available.

7.3 Chewing recognition

For the recognition of chewing sounds, novel achievements on a chew-accurate detection were presented. Using the recognition procedure, individual chewing cycles were identified in two food categories with good performance. This result was achieved by considering the chew as a non-stationary event and grouping the foods with similar textures. In comparison to our earlier investigation ([29]), the current recognition rates are approx. 15% higher and a majority vote over multiple chewing cycles could be avoided. However, for low-amplitude chewing sounds, found in soft foods such as cooked pasta or meat, a low detection performance persists with the current approach. This effect was attributed to the low signal to noise ratio of these sounds. Moreover, the chewing sequence is not consistent over the entire intake cycle as assumed in the current approach [42]. This is observed as a variability in the detection confidences and hinders fusion methods such as LR to achieve a higher performance. Consequently, food models should include the sequence information more carefully.

7.4 Swallowing recognition

The automatic detection of swallowing using EMG and sound information was evaluated. We found that swallows can be retrieved from continuous data at high recall rates using both sensing sources. By observing the final detection, we found that the method is disturbed by neck movements and coughing. In comparison to our previous work ([38]), we presented results from additional fusion methods (AGREE, LR) and an extended study. The AGREE fusion was able to remove a large share of insertion errors. The current results confirm the previous findings: while the detection works to some extend in controlled environments, it retrieved many false positives in our evaluation. These errors could not be completely removed by the currently applied fusion techniques.

The collar worked well to standardise and maintain the sensor positioning. No differences in the spotting results were observed for the collar-based swallowing data. For a subgroup of two participants an improved performance was achieved. The difference could not be explained by the available information.

A larger studies could reveal, whether the subgroups persist. Further investigations are required to analyse options for food bolus categorisation and to increase the algorithm precision.

8 Conclusion

We presented novel approaches to monitor dietary activities from body-worn sensors. Three sensing domains were analysed, that are directly linked to the sequence of dietary activities: intake movements, chewing, and swallowing. We presented evaluation results from studies in each domain using an event recognition procedure, that supports the detection and identification of specific activities in continuous sensor data.

The recognition of natural movements, such as for dietary intake, is a challenging task, since it is strongly related to personal habits. The detection procedure in combination with the simple comparison fusion yielded good recognition results for different intake types. This is a valuable result for the intended application, since the intake movements help to categorise the consumed foods. Moreover, the movement recognition could be used independently. For example, the detection of drinking movements can be used to monitor fluid consumption and avoid dehydration.

Chewing is a very important part in the intake process. In this work a successful continuous recognition of two food types was achieved. This is a vital result for a detailed analysis of food chewing. Based on the presented approach, additional models can be derived that reflect the mechanical properties of foods. Besides the identification of consumed foods, the chewing recognition permits the assessment of dietary parameters, such as chews per food and chewing speed. Both parameters can be used as indications for too fast, or stress eating.

Swallowing concludes the intake cycle. The swallowing frequency depends on the food category, where foods containing fluid compartments require elevated swallowing rates. The current detection method, using sound and muscle activity at the throat, still incurs many insertion errors. However, it does provides an indication for swallowing events. We plan to use this information in combination with the previous sensing domains. Further works will address different fusion strategies and additional sensors.

The three domains provide a comprehensive picture of dietary activities and a broad amount of information, that is vital for a long-term dietary coaching and health management. This includes the food type as well as intake timing and the overall meal schedule.

We have shown in this work, how our recognition procedure to spot sporadic activity events can be slightly adapted to fulfil the requirements of very different sensor modalities and activities. We believe that the procedure is a helpful tool for Automatic Dietary Monitoring and similar applications in continuous activity recognition.

Acknowledgements

The authors express their gratitude to all volunteers who participated in the studies related to this publication and to all reviewers for their very helpful comments. This work was supported by the Swiss State Secretariat for Education and Research (SER).

References

[1] WHO, Global strategy on diet, physical activity and health (WHA57.17), in: Fiftyseventh World Health Assembly, World Health Organization, 2004.

[2] A. Sjöberg, L. Hallberg, D. Höglund, L. Hulthen, Meal pattern, food choice, nutrient intake and lifestyle factors in the Göteborg Adolescence Study., European Journal of Clinical Nutrition 57 (12) (2003) 1569–1578.

[3] B. J. Rolls, A. Drewnowski, J. H. Ledikwe, Changing the energy density of the diet as a strategy for weight management., Journal of the American Dietetic Association 105 (5 Suppl 1) (2005) S98–103.

[4] MyFoodPhone, World's first camera-phone & web-based-video nutrition service, Internet, accessed: August 2007 (Feb 2005).

[5] J. Beidler, A. Insogna, N. Cappobianco, Y. Bi, M. Borja, The PNA project, Journal of Computing Sciences in Colleges 16 (4) (2001) 276–284.

[6] J. Mankoff, G. Hsieh, H. C. Hung, S. Lee, E. Nitao, Using low-cost sensing to support nutritional awareness, in: G. Goos, J. Hartmanis, J. van Leeuwen (Eds.), Ubicomp 2002: Proceedings of the 4th International Conference on Ubiquitous Computing, Vol. 2498 of Lecture Notes in Computer Science, Springer Berlin, Heidelberg, 2002, pp. 371–376.

[7] K. A. Siek, K. H. Connelly, Y. Rogers, P. Rohwer, D. Lambert, J. L. Welch, When do we eat? an evaluation of food items input into an electronic food monitoring application, in: E. Aarts, R. Kohno, P. Lukowicz, J. C. Trainini (Eds.), PHC 2006: Proceedings of the 1st International Conference on Pervasive Computing Technologies for Healthcare, ICST, IEEE digital library, 2006, pp. 1–10.

[8] A. A. Stone, S. Shiffman, J. E. Schwartz, J. E. Broderick, M. R. Hufford, Patient non-compliance with paper diaries., British Medical Journal 324 (7347) (2002) 1193–1194.

[9] M. S. Westerterp-Plantenga, Eating behavior in humans, characterized by cumulative food intake curves–a review, Neuroscience and Biobehavioral Reviews 24 (2) (2000) 239–248.

[10] H. R. Kissileff, G. Klingsberg, T. B. V. Itallie, Universal eating monitor for continuous recording of solid or liquid consumption in man., The American Journal of Physiology 238 (1) (1980) R14–R22.

[11] E. Stellar, E. E. Shrager, Chews and swallows and the microstructure of eating., The American Journal of Clinical Nutrition 42 (5 Suppl) (1985) 973–982.

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

[13] A. Schmidt, M. Strohbach, K. van Laerhoven, A. Friday, H.-W. Gellersen, Context acquisition based on load sensing, in: G. Goos, J. Hartmanis, J. van Leeuwen (Eds.), Ubicomp 2002: Proceedings of the 4th international conference on Ubiquitous Computing, Vol. 2498 of Lecture Notes in Computer Science, Springer Berlin, Heidelberg, 2002, pp. 333–350.

[14] M. Beigl, H.-W. Gellersen, A. Schmidt, MediaCups: Experience with design and use of computer-augmented everyday artefacts, Computer Networks 35 (4) (2001) 401–409, special Issue on Pervasive Computing.

[15] D. Patterson, D. Fox, H. Kautz, M. Philipose, Fine-grained activity recognition by aggregating abstract object usage, in: B. Rhodes, K. Mase (Eds.), ISWC 2005: Proceedings of the Ninth IEEE International Symposium on Wearable Computers, IEEE Press, 2005, pp. 44–51.

[16] S. Chambers, S. Venkatesh, G. West, H. Bui, Hierarchical recognition of intentional human gestures for sports video annotation, in: R. Kasturi, D. Laurendeau, C. Suen (Eds.), Proceedings of the 16th International Conference on Pattern Recognition, Vol. 2, IEEE Press, 2002, pp. 1082–1085.

[17] G. Ogris, T. Stiefmeier, H. Junker, P. Lukowicz, G. Troster, Using ultrasonic hand tracking to augment motion analysis based recognition of manipulative gestures, in: B. Rhodes, K. Mase (Eds.), ISWC 2005: Proceedings of the Ninth IEEE International Symposium on Wearable Computers, IEEE Press, 2005, pp. 152–159.

[18] H.-K. Lee, J. H. Kim, An HMM-based threshold model approach for gesture recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (10) (1999) 961–973.

[19] J. Ward, P. Lukowicz, G. Tröster, T. Starner, Activity recognition of assembly tasks using body-worn microphones and accelerometers, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (10) (2006) 1553–1567.

[20] C. Lee, X. Yangsheng, Online, interactive learning of gestures for human/robot interfaces, in: N. Caplan, C. G. Lee (Eds.), ICRA 1996: Proceedings of the IEEE International Conference on Robotics and Automation, Vol. 4 of IEEE Robotics and Automation Society, IEEE Press, 1996, pp. 2982–2987.

[21] O. Amft, H. Junker, G. Tröster, Detection of eating and drinking arm gestures using inertial body-worn sensors, in: B. Rhodes, K. Mase (Eds.), ISWC 2005: IEEE Proceedings of the Ninth International Symposium on Wearable Computers., IEEE Press, 2005, pp. 160–163.

[22] Z. M. Vickers, The relationships of pitch, loudness and eating technique to judgments of the crispness and crunchiness of food sounds, Journal of Texture Studies 16 (1) (1985) 85–95.

[23] C. Dacremont, B. Colas, F. Sauvageot, Contribution of air- and bone-conduction to the creation of sounds perceived during sensory evaluation of foods, Journal of Texture Studies 22 (4) (1991) 443–456.

[24] W. AlChakra, K. Allaf, A. Jemai, Characterization of brittle food products: Application of the acoustical emission method, Journal of Texture Studies 27 (3) (1996) 327–348.

[25] J. Edmister, Z. Vickers, Instrumental acoustical measures of crispness in foods, Journal of Texture Studies 16 (2) (1985) 153–167.

[26] B. Drake, Food crushing sounds. an introductory study, Journal of Food Science 28 (2) (1963) 233–241.

[27] N. DeBelie, V. De Smedt, D. B. J., Principal component analysis of chewing sounds to detect differences in apple crispness, Journal of Postharvest Biology and Technology 18 (2000) 109–119.

[28] N. DeBelie, M. Sivertsvik, J. DeBaerdemaeker, Differences in chewing sounds of dry-crisp snacks by multivariate data analysis, Journal of Sound and Vibration 266 (3) (2003) 625–643.

[29] O. Amft, M. Stäger, P. Lukowicz, G. Tröster, Analysis of chewing sounds for dietary monitoring, in: M. Beigl, S. Intille, J. Rekimoto, H. Tokuda (Eds.), UbiComp 2005: Proceedings of the 7th International Conference on Ubiquitous Computing., Vol. 3660 of Lecture Notes in Computer Science, Springer Berlin, Heidelberg, 2005, pp. 56–72.

[30] C. Ertekin, I. Aydogdu, Neurophysiology of swallowing., Clinical Neurophysiology 114 (12) (2003) 2226–2244.

[31] D. M. Denk, H. Swoboda, E. Steiner, Physiology of the larynx, Der Radiologe 38 (2) (1998) 63–70, in German.

[32] V. Gupta, N. P. Reddy, E. P. Canilang, Surface EMG measurements at the throat during dry and wet swallowing., Dysphagia 11 (3) (1996) 173–179.

[33] W. J. Logan, J. F. Kavanagh, A. W. Wornall, Sonic correlates of human deglutition., Journal of Applied Physiology 23 (2) (1967) 279–284.

[34] J. A. Y. Cichero, B. E. Murdoch, Detection of swallowing sounds: methodology revisited., Dysphagia 17 (1) (2002) 40–49.

[35] C. Danbolt, P. Hult, L. T. Grahn, P. Ask, Validation and characterization of the computerized laryngeal analyzer (CLA) technique., Dysphagia 14 (4) (1999) 191–195.

[36] A. Limdi, M. McCutcheon, E. Taub, W. Whitehead, I. Cook, E.W., Design of a microcontroller-based device for deglutition detection and biofeedback, in: EMBS 1989: Proceedings of the Annual International Conference of the IEEE Engineering in Engineering in Medicine and Biology Society., Vol. 5, IEEE Press, 1989, pp. 1393–1394.

[37] S. M. Sukthankar, N. P. Reddy, E. P. Canilang, L. Stephenson, R. Thomas, Design and development of portable biofeedback systems for use in oral dysphagia rehabilitation., Medical Engineering & Physics 16 (5) (1994) 430–435.

[38] O. Amft, G. Tröster, Methods for detection and classification of normal swallowing from muscle activation and sound, in: E. Aarts, R. Kohno, P. Lukowicz, J. C. Trainini (Eds.), PHC 2006: Proceedings of the First International Conference on Pervasive Computing Technologies for Healthcare, ICST, IEEE digital library, 2006, pp. 1–10.

[39] T. K. Ho, J. Hull, S. Srihari, Decision combination in multiple classifier systems, IEEE Transactions on Pattern Analysis and Machine Intelligence 16 (1) (1994) 66–75.

[40] D. Bannach, O. Amft, K. S. Kunze, E. A. Heinz, G. Tröster, P. Lukowicz, Waving real hand gestures recorded by wearable motion sensors to a virtual car and driver in a mixed-reality parking game, in: A. Blair, S.-B. Cho, S. M. Lucas (Eds.), CIG 2007: Proceedings of the 2nd IEEE Symposium on Computational Intelligence and Games, IEEE Press, 2007, pp. 32–39.

[41] J. A. Ward, P. Lukowicz, G. Tröster, Evaluating performance in continuous context recognition using event-driven error characterisation, in: M. Hazas, J. Krumm, T. Strang (Eds.), LoCA 2006: Proceedings of the Second International Workshop on Location- and Context-Awareness, Vol. 3987 of Lecture Notes in Computer Science, Springer, Berlin/Heidelberg, 2006, pp. 239–255.

[42] O. Amft, M. Kusserow, G. Tröster, Automatic identification of temporal sequences in chewing sounds, in: T. Hu, I. Mandoiu, Z. Obradovic (Eds.), BIBM2007: Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine, IEEE Press, San Jose, CA, USA, 2007, pp. 194–201.