

Methods for Detection and Classification of Normal Swallowing from Muscle Activation and Sound

Oliver Amft and Gerhard Tröster

Abstract—Swallowing is an important part of the dietary process. This paper presents an investigation to detect and classify normal swallowing during eating and drinking from electromyography and microphone sensors. The non-invasive sensors are selected in order to integrate them into a collar-like fabric for continuous monitoring of swallowing activity over a day. We compare methods for the detection of individual swallowing events from continuous sensor data. Furthermore we present a classifier comparison for the swallowing event properties volume and viscosity. The methods are evaluated on experimental data and a performance analysis is shown. Moreover we present a class skew analysis based on the metrics precision and recall.

Index Terms—Swallowing detection, event detection, bolus viscosity classification, bolus volume classification, sensor collar.

I. INTRODUCTION

The prevalence of chronic diseases related to lifestyle and behaviour as well as the aging population leads to a surge of healthcare costs all over the world. Consequently new concepts and methods are needed to fight diseases such as obesity, hypertension and cardio-vascular diseases. It is envisioned that long-term behavioural monitoring and coaching can contribute vastly to the problem of maintaining or achieving a healthy lifestyle and therefore reducing the risks of these diseases.

Relevant lifestyle aspects related to the afore-mentioned diseases include exercise and dietary behaviour. Our work aims at developing methods to monitor dietary behaviour automatically. We believe that wearable systems can provide valuable insight into daily eating behaviour, that is difficult to achieve by other means. The work on swallowing detection presented in this paper is considered one part of a wearable dietary monitoring system, since swallowing is inherently linked to eating and drinking activities.

A. Automatic dietary monitoring

Dietary monitoring includes a variety of aspects such as timing and frequency of eating activities, rate of intake as well as type and amount of foodstuff. Information about these parameters on a daily basis provide insight into the dietary activities and can be integrated in lifestyle feedback and reminders that have a relevant impact, e.g. to maintain a lunch duration of at least 15 minutes. Currently dietary activities are studied exemplary by entering the information manually into questionnaires. This involves a considerable effort of study participants and managers.

We believe that the absolute error-free estimation of amount and calories of every possible nutrient is rather visionary, using

non-invasive sensors. However, a rough estimation of food type, e.g. ratio of fluid and solid nutrient combined with the timing information, e.g. event schedule and meal durations over the day, already provides a solid basis for behavioural monitoring. Although focusing on wearable sensors we expect that additional information can be obtained in combination with a supportive environment, e.g. food products with RF-identification tags, intelligent shopping lists or dietary monitoring tables.

We target a non-invasive wearable system relying on information from the following three sensing domains: 1) the identification of characteristic arm and trunk movements associated with food intake using inertial sensors [1], 2) the analysis of food chewing sounds from an ear microphone [2] and 3) the detection of swallowing from body-worn sensors. The focus of this paper is on the latter.

B. Swallowing process

Swallowing is a frequent human activity. It is estimated that normal swallowing occurs approx. 600 to 2000 times per day in healthy persons [3].

The swallowing act is often partitioned into three distinctive phases [3]: 1) the oral preparation, 2) the pharyngeal, and finally 3) the esophageal phase. During the oral phase a food piece is transformed to a swallowable bolus. This may involve chewing and forming a bolus by tongue movements (depending on the food texture) and initiating the swallowing reflex, which starts the pharyngeal phase. In the oral phase the bolus type is sensed with regard to volume and viscosity. Henceforth the swallowing apparatus may adapt to the bolus [4].

The pharyngeal phase is formed by the bolus travelling through the pharynx and passing the upper esophageal sphincter. During this phase a sequence of muscle activations is used to propel the bolus and protect the trachea from contamination. The following esophageal phase is composed of peristalsis contractions that move the bolus towards the stomach.

Since the oral phase is involved with the variable process of chewing, it is less informative for the detection of swallowing. The pharyngeal and esophageal phase are expected to be more specific since these are not controlled voluntarily. However the latter cannot be accessed with non-invasive methods due to the spine and trachea covering the esophagus. Hence the pharyngeal swallowing phase is addressed with non-invasive sensors.

C. Paper contributions

The work presented in this paper aims at utilising non-invasive sensing modalities to detect and identify swallowing at the pharyngeal phase. Specifically, the following contributions are made:

- 1) We propose sensor modalities and locations that support the identification of individual swallows and present a experimental methodology to evaluate the feasibility of these sensor types during daily activities. Moreover we address the restricted sensor positioning that stems from the goal to integrate the sensors into a collar-like fabric.
- 2) We compare the performance of two swallowing event detection approaches on continuous sensor data. Here, the goal is to separate the swallowing events from sensor noise incurred from everyday activities and the various other functions of the pharynx.
- 3) We evaluate classifiers for the discrimination of bolus volume and viscosity and present classification results that indicate the discriminative information extracted from the chosen sensor modalities.

The work presented here is a first attempt to detect and classify swallowing events automatically and evaluate different procedures. The envisioned detection system shall not hinder the user's perception and a deployment in non-clinical environments is aimed. Specifically we rely on surface electromyography (SEMG) detection of muscle activation patterns and sounds associated with the swallowing event. We evaluate the different classification and event detection algorithms on recordings from 5 subjects and a total of 868 annotated swallows.

D. Related Work

A number of clinical assessment methods have been developed to analyse the complex interaction of swallowing with phonation and respiration at the throat level. The most important invasive methods include videofluoroscopy, e.g. [5], manometry, e.g. [6] and wire-electrode based electromyography (EMG), e.g. [7].

A number of non-invasive assessment methods have been evaluated during pharyngeal swallowing, including sensing of muscle activations by SEMG, e.g. [8], [9], listening to the throat sounds (cervical auscultation) by stethoscope [10] and stethoscope acoustic transducers or sealed microphones [11]. As alternative to the acoustic analysis, tissue vibrations have been analysed [12]. However no clear advantage of the vibration based analysis was reported, except that the vibration sensor is more robust against environmental noises at the expense of a much higher device cost.

Some works aimed at sensing the larynx movement by using movement sensors at the neck, e.g. [13]. However the detection performance is strongly depending on gender with weak results at the less prominent female larynx. Furthermore it was shown that the simple sensor incurs errors from neck and tongue movements as well as larynx movements during speaking or externally applied pressure when used during daily activities.

Several other approaches have been proposed for the analysis of swallowing, mostly in combination with previously mentioned invasive methods, including ultrasound [4], [14], pharyngeal impedance sensing methods [15]–[18] and impedance plethsmography [19].

Different automatic feedback systems for the detection of swallowing abnormalities such as dysphagia have been proposed. Most of the abnormality detection approaches rely on SEMG or vibration sensors, e.g. [20]. These systems classify the subjects based on isolated swallowing events that have been identified and marked manually by an expert.

A few attempts have been made to detect swallowing from continuous sensor readings. Pehlivan et al. [21] proposed a device for counting swallows and compared swallow counts in normal subjects and Parkinson's disease subjects during eating and drinking. The device is based on the mechanical sensing approach using a piezoelectric sensor attached to the larynx. A manual pre-segmentation for nutrition phases was applied. Speech was specifically excluded.

Das et al. [22] deployed an ensemble of neural networks to discriminate normal and dysphagic swallows from vibration recordings. In their approach swallows were recorded in a controlled environment largely avoiding sensor artefacts. Persisting artefacts were segmented by modelling them specifically with the neural networks.

The approach of Limdi et al. [23] was based on SEMG intensity detection and aimed at informing the user of elevated swallowing rates. Sukthakar, Reddy et al. [24] used SEMG and vibration sensors and aimed at dysphagia rehabilitation. However both works did not present a performance evaluation of their approaches for the continuous detection problem.

The pharyngeal phase of swallowing is influenced by the type of swallowed foodstuff. During chewing and tongue movement the bolus is sensed by various receptors in the oral cavity. Specifically volume, mass and viscosity of the bolus modify the central neurological pattern generator [4], [7]. Dantas et al. [25] found that bolus transit time through the pharynx increases with viscosity. This effect was captured in duration and amplitude parameters of SEMG recordings from the submental and infra-hyoid regions [25]–[27]. Moreover it was found in these studies that transit time and SEMG features are largely unaffected by bolus volume.

Chichero et al. [28] and Boiron et al. [29] reported a dependency of swallow sound features on bolus volume. However the studies disagree on the type of interaction.

E. Paper structure

The remaining of the paper is structured as follows: Section II describes our event detection and classification approach as well as the conducted experiments. Section III summarises the results of the swallowing event detection from continuous data. Section IV provides the results of the bolus type classification. Finally, Section V provides a conclusion followed by an outlook on further work.

II. METHODOLOGY

This section provides an overview on our approach to detect and classify swallowing. Furthermore the experimental protocol to acquire evaluation data is described.

A. Approach

As described in the introduction of this paper our detection and classification targets the analysis of pharyngeal swallowing using non-invasive sensors attached to the user's neck. Fig. 1 illustrates the overall concept of our approach to the problems of sensor data acquisition, event detection and classification. The following sections of this paper will evaluate solutions for these problems.

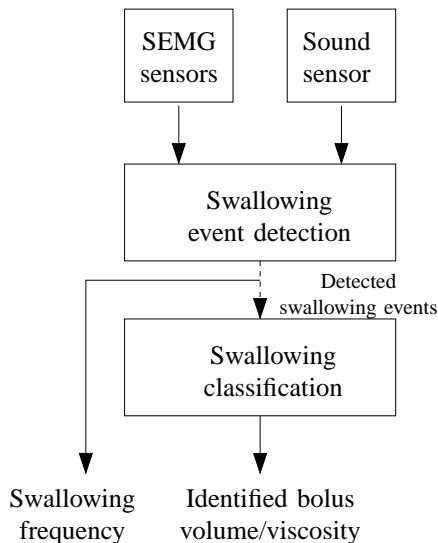


Fig. 1. Approach to search and identify swallowing events. (In this paper the swallowing classification is supported by manual annotation information.)

Swallowing data acquisition is related to the problem of selecting appropriate non-invasive sensors that provide means for extracting information on swallowing events and the bolus characteristics viscosity and volume from the pharyngeal swallowing phase. Following the findings in [26] we recorded SEMG to capture the viscosity variability and sound to analyse the volume/density variability of the pharynx [28]. Details of the experimental procedure and sensor placement are described in the following Section II-B.

The swallowing event detection aims at extracting signal sections that contain individual swallows from a continuous stream of sensor data. Specifically the challenge is to distinguish swallowing events from sensor noise and artefacts, recorded when wearing the system during daily activities. By selecting a experimental procedure that includes non-swallowing activities, e.g. speaking, head turning, chewing, we aimed to cover these situations.

We evaluate two different methods for the swallowing event detection: 1) using a simple signal intensity measure applied to the rectified SEMG amplitude and 2) using a pattern search based on a similarity measure of a data section. Here the pattern of a swallowing event is described using features derived

from the sensor data. Section III presents the procedures and the evaluation results in detail.

We analyse the feasibility to discriminate bolus viscosity and volume using the sensor data in Section IV. Our approach is based on an isolated classification of individual voluntary swallows. For the investigation in this paper a manual annotation was applied to isolate the swallows. Fused SEMG and sound feature sets from time-domain and combined frequency- and time-domain were evaluated by analysing the classification performances.

B. Experiments

1) *Test subjects and materials:* Five subjects (3 male, 2 female, aged 20 to 30 years) without known swallowing abnormalities were instructed to eat and drink different food-stuff items: 5 and 15 ml of water, a spoonful of yoghurt and a piece of bread (approx. 2 cm³). The items are summarised in Tab. I. The material size was controlled by syringe for the water and visually for the spoonful of yoghurt and the bread pieces. Additionally, reference samples from all foods were weighted.

The subjects were asked to aim at swallowing the nutrient items in one piece after chewing and manipulating the bolus as usual. None of the subjects expressed a dislike for any of the covered nutrients nor problems to swallow the selected bolus sizes. Subjects were sitting conveniently on a chair close to a table carrying the nutrients. They were allowed to move, chew and speak normally during the recording sessions. Naturally, during the short pharyngeal swallowing phases on speaking was audible. The environment was controlled for low and constant noise level during the swallowing events.

2) *Sensor selection and location:* Surface EMG from submental (SM-EMG) and infra-hyoid (IH-EMG) regions were recorded by gel electrodes at 24 bit, 2 kHz and bandpass filtered. Swallowing sound was recorded by an electret condenser microphone (type Sony ECM-C115), placed inferior midline from the cricoid cartilage. The microphone was secured and sealed with medical tape, following the protocol of previous investigations [11], [12], [28]. Sound data was recorded at 16 bit, 22 kHz. For the individual analysis steps the sample rate of SEMG and sound was reduced.

Fig. 2 illustrates the positioning of the sensors. These positions have been used by previous investigations on SEMG [26] and sound [28]. The SM-EMG electrode set was included in the recordings mainly for comparison and swallowing event inspection purposes.

3) *Recording procedure:* The nutrient properties are listed in Tab. I. The subjects were instructed to eat/drink items from each of the nutrient categories to obtain at least 15 swallows per session. To account for physiologic variations two sessions were recorded on different days. The recording duration was not constraint since the subject were eating/drinking at their individual speeds, selecting the food category for each individual swallow.

An observer was verifying the procedure during each session and annotating the food category as well as begin/end of each swallowing event. Additionally all recording sessions

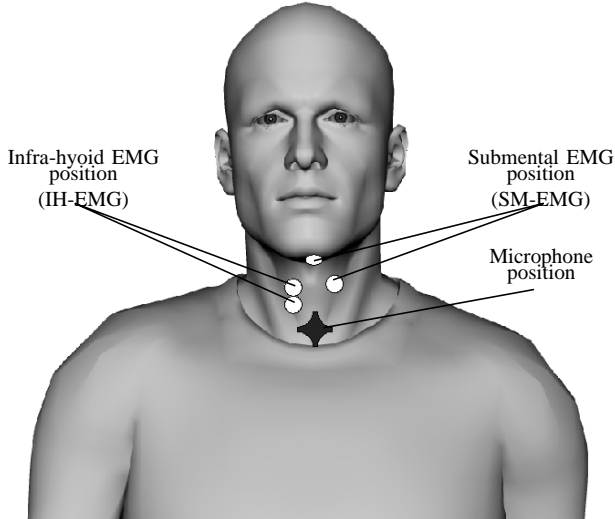


Fig. 2. Schematic sensor positioning at the neck.

were videotaped for later verification of the annotated events. To simplify the online annotation, subjects were instructed to indicate swallowing to an observer by raising the hand and stop chewing shortly before swallowing. In a post-processing step all annotated events were reviewed and the begin/end times were adapted by the observer inspecting the signals. In situations where the swallowing event could not be clearly identified, the sound data was played back and/or the recorded video was analysed. In some situations spontaneous swallowing or multiple swallowing occurred before/after swallowing the food item. It was assumed that these swallows were used to clear the oral cavity and resulted in a tiny bolus or saliva swallow. These swallows were annotated as 5 ml water swallows.

To achieve a data level alignment artificial synchronisation events have been inserted on both data streams (SEMG and sound) during the recordings. In the post-processing step the synchronisation events were used to adapt the alignment of the data streams.

TABLE II
SUMMARY OF EVALUATION DATA FROM 5 SUBJECTS.

Total swallowing events	868
Total swallowing duration	1632 s (27.2 min)
Total length of dataset	17465 s (4.85 hours)

Tab. II summarises the recorded and inspected swallowing events. To obtain a realistic data set additional data, resembling daily activities, were recorded with all subjects wearing the sensors. These activities included, speaking and conversation, background noise, head turning, tilting, nodding and chewing. In total 868 swallowing events were recorded and inspected from 4.85 hours of sensor data.

III. DETECTION OF SWALLOWING EVENTS

In order to analyse and classify individual swallows, data sections containing the swallowing events need to be extracted from the continuous stream of sensor data.

The challenge to detect swallowing events can be formulated as follows: the envisioned system shall be continuously worn during daily activities, however swallowing events occur comparably rarely, embedded in non-swallowing phases (NULL class). A method aiming at detecting swallows shall be effective in retrieving correct events and omitting non-swallow phases while maintaining a low processing effort. The approach presented here attempts to isolate swallowing events for later analysis. Consequently the methods are optimised to reduce event misses at the expense of increased false positives. Moreover, the swallowing phases have a variable length as the event durations in Tab. I indicate.

A. Signal intensity detection

EMG signals are usually rectified and averaged for human inspection. In this way muscle contractions can be spotted visually as peaks in the waveform. We utilise a similar approach for detecting muscle contractions during swallowing events automatically: by sweeping a threshold on the rectified EMG amplitude possible events are obtained as signal sections, where the amplitude is above the threshold. Selecting the threshold controls the system performance, e.g. the rate of false positives and false negatives.

We used the IH-EMG data at a resolution of 256 Hz. The rectified IH-EMG was obtained by averaging the absolute signal amplitude using a sliding window of 32 samples, one sample step size.

Since this method is not sensitive to the data pattern of a swallowing event, except for the signal intensity, it incurs errors and can be used to qualify the evaluation data. For the IH-EMG intensity, more detection errors correspond to more sensor artefacts from chewing and other pharyngeal activities. Furthermore the thyrohyoid muscle targeted with the infra-hyoid surface electrode position is covered by other muscle layers, disturbing the activation detection.

B. Feature similarity detection

The feature similarity approach relies on a two-step procedure of signal segmentation and similarity search. In the first step segmentation points are determined that reduce the subsequent search effort for the similarity analysis. We used the Sliding-Window And Bottom-up (SWAB) algorithm [30]. This algorithm partitions a continuous stream of sensor data very robustly by sequentially testing the approximation of the signal through linear regression lines and using the boundaries of these approximations as segments. The duration of the signal segments describe the signal variation over time, with shorter segments for highly fluctuating signals and longer segments for relatively monotone phases. We applied this algorithm to the rectified averaged IH-EMG signal (256 Hz signal resolution, mean window size of 32 samples, one sample step size). Fig. 3 illustrates the obtained segmentation boundaries at a sample signal.

The second step utilises the IH-EMG segmentation points to search for swallowing event sections using a feature similarity measure. The search is performed by analysing the similarity of a data section under investigation compared to a trained

TABLE I
DESCRIPTION OF EVALUATION DATA: NUTRIENTS AND PROPERTIES.

Foodstuff	Item volume	Viscosity category	Total swallowing events	Annotation duration (mean \pm SD)	SM-EMG activation ^a (mean \pm SD)	IH-EMG activation ^a (mean \pm SD)
Water	5 ml	fluid	353	1.8 s \pm 0.5 s	0.49 \pm 0.13	0.55 \pm 0.17
Water	15 ml	fluid	205	2.0 s \pm 0.5 s	0.55 \pm 0.13	0.59 \pm 0.16
Yoghurt	\sim 7 ml (filled spoon)	semifluid	171	1.8 s \pm 0.5 s	0.54 \pm 0.14	0.56 \pm 0.15
Bread	\sim 2 cm ³ (prepared pieces)	solid	139	2.0 s \pm 0.6 s	0.54 \pm 0.14	0.64 \pm 0.17

^aRectified SEMG signal duration above minimum+1 SD within the annotation section.

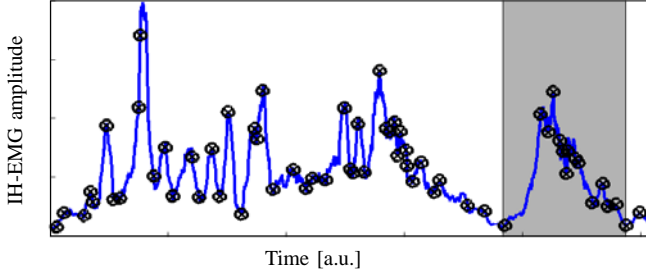


Fig. 3. Rectified IH-EMG sample signal of yoghurt eating (mean-filtered). The SWAB segmentation points are marked with \otimes . The shaded section indicates an annotated swallow.

pattern. For a given segmentation point, the history of sensor data is analysed from a lower up to an upper search bound. These bounds are determined in the training step from minimum/maximum overlaps between the annotated events and the segmentation points.

The similarity of the sensor data is determined from the Euclidean distance between the features of the data section under investigation during the search and the trained pattern. This approach has been applied previously to a classification problem of movement data from inertial sensors [1], [31].

The results of the feature similarity search is a list of data sections with an associated Euclidean distance. From the training data an optimal distance threshold is determined that retains the best matching sections with the manual annotation.

The feature similarity search procedure was applied to features from IH-EMG and sound data individually and combined using feature-level fusion. The features from feature set 1 (time domain, see Tab. V) were used for the evaluation of the similarity searches. With regard to the potentially low mobile processing performance a low data resolution was used for both IH-EMG (128 Hz) and sound (4 kHz).

Furthermore two event fusion methods were tested: 1) a comparison of the individual IH-EMG and sound event detections and 2) a second-pass similarity search.

The comparison of sensor-specific event detections aims at selecting the front of best events from the individual similarity searches. For this procedure a detection confidence was determined by normalising the sensor-specific event distances with the corresponding similarity training threshold. In this way the event detection results of independent similarity searches can be compared. The best events were selected by a sliding

window procedure.

For the fusion using a second-pass similarity search we applied an additional training step based on the event confidences of the individual similarity results. The training data from the first-pass similarity was reused for this training. The confidence was determined in the same way as for the comparison method.

C. Evaluation procedure

Training and testing was performed on the subject-specific data sets. To account for variations in the data set a 4-fold cross-validation procedure was used to determine training and testing data set for both detection procedures, IH-EMG intensity and feature similarity search. For the training 3 of 4 data parts were used. Evaluation was performed on the left out data part. This procedure was repeated until all 4 parts were used for testing once. The partition boundaries were adapted to avoid intersecting swallowing data sections.

To analyse performance, we utilised the metrics *Precision* and *Recall* commonly used for evaluation in Information Retrieval. These metrics are derived as follows:

$$Recall = \frac{TP}{P} = \frac{\text{Recognised swallows}}{\text{Relevant swallows}} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} = \frac{\text{Recognised swallows}}{\text{Retrieved swallows}} \quad (2)$$

Relevant swallows corresponds to the manually annotated number of swallowing events in a class (positives, P). Retrieved swallows represents the number of swallowing events that are returned by the algorithm. This includes both, true positives (TP) and false positives (FP). Finally, recognised swallows refers to the correctly returned number of swallowing events (true positives, TP).

D. Detection results

The results of all investigated detection methods are summarised in Tab. III: IH-EMG intensity (Intensity), feature similarity (SIM) for IH-EMG, sound and the feature level fusion, as well as the event fusion methods comparison (COMP) and second-pass. For all methods a threshold was chosen to achieve high recall. For the comparison both recall and precision must be considered.

TABLE III
SUMMARY FOR THE SUBJECT-SPECIFIC DETECTION PERFORMANCE.

Sensors	Intensity	SIM	SIM	SIM	COMP	2nd pass
	IH-EMG	IH-EMG	SND	IH-EMG &SND	IH-EMG &SND	IH-EMG &SND
Relevant	868	868	868	868	868	868
Retrieved	8065	4128	4368	3961	1853	1660
Recognised	645	715	634	726	567	491
FN	223	153	234	142	301	377
FP	7420	3413	3734	3235	1286	1169
Recall	0.74	0.82	0.73	0.84	0.65	0.57
Precision	0.08	0.17	0.15	0.18	0.31	0.30

1) *IH-EMG intensity*: An overall recall of 0.74 was achieved on the evaluation data. However the method retrieves many false positives (low precision value). The weak detection result can be accounted to the fluctuating signal with high amplitude values for arbitrary muscle contractions and artefacts.

2) *Feature similarity*: The SWAB algorithm obtained 76803 segments for the 4.85 hours of evaluation data. The jitter between segmentation boundary and manual swallowing event annotations were analysed for all food categories. The mean jitter was below 0.12 s (SD: 0.11 s) for all 868 events.

An overall recall of 0.82, 0.73 and 0.84 was achieved for IH-EMG, sound and the feature-level fusion respectively. However the feature similarity searches retrieved far less false positive errors (higher precision value), compared to the IH-EMG intensity method. This is illustrated in the threshold sweep of Fig. 4. While the intensity method reaches an acceptable recall level, the precision does not increase above 0.1.

The precision-recall comparison in Fig. 4 furthermore presents the modelling performance of the two event fusion methods: similarity comparison (COMP) and second-pass similarity. While both methods retrieve far less false positives (increased precision), the number of recognised swallowing events decrease, when compared to the sensor-specific similarity searches. In the depicted example, the second-pass similarity provides (as intended) a far better result, resembling the performance of the best sensor-specific similarity result while the comparison method incurs more recognition errors.

IV. CLASSIFICATION OF SWALLOWING EVENTS

As reviewed in the introduction, different interactions of bolus volume and viscosity with SEMG and sound features have been tested in the clinical settings of previous studies. In this section we evaluate the following hypotheses: 1) IH-EMG supports the discrimination of the bolus viscosity independent from volume and 2) sound supports the discrimination of the bolus volume. Our analysis is based on the isolated classification using manually derived swallowing annotation.

A. Evaluation procedure

In order to investigate the hypotheses described above the recorded food categories were grouped into classes according to Tab. IV. We assumed here that the chosen foodstuffs represent the typical variations in foods with regard to viscosity (fluid, semifluid and non-fluid) as well as volume (see Tab. I).

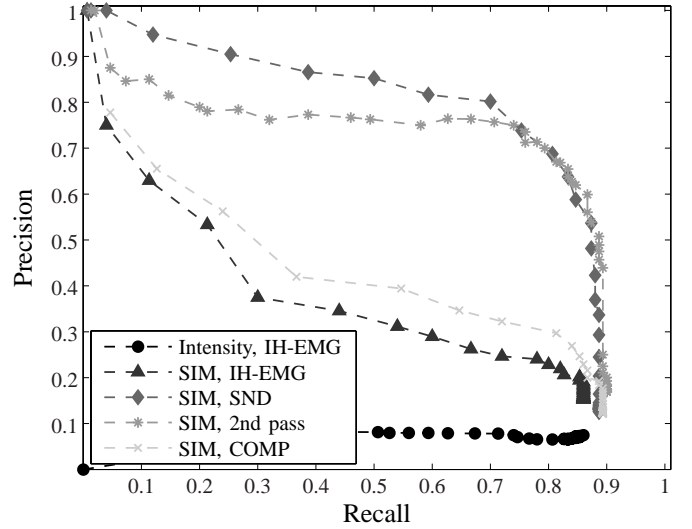


Fig. 4. Precision-recall comparison (threshold sweep) of the IH-EMG intensity and feature similarity search methods (SIM) of a trained data section containing 150 relevant swallows from one subject. Best performance is found towards the top-right corner (high precision, high recall).

TABLE IV
CLASS GROUPINGS OF NUTRIENT CATEGORIES.

Hypothesis		Food group classes
Volume	1	Class 1: 5 ml water, 2 cm ³ bread pieces Class 2: spoonful yoghurt Class 3: 15 ml water
	2	Class 1: 2 cm ³ bread pieces, 5 ml water, spoonful yoghurt Class 2: 15 ml water
Viscosity	1	Class 1: 5 ml+15 ml water Class 2: spoonful yoghurt Class 3: 2 cm ³ bread pieces
	2	Class 1: 5 ml+15 ml water Class 2: spoonful yoghurt, 2 cm ³ bread pieces

Two feature sets (summarised in Tab. V) were computed from the sensor data. Feature set 1 is based on time domain properties of the sensor streams, feature set 2 contains set 1 and additional frequency domain features. Feature set 1 was processed at relatively low sampling resolution of 128 Hz for SEMG and 4 kHz for sound. For feature set 2, higher frequencies were used: 2 kHz for SEMG and 16 kHz for the sound.

The feature sets were evaluated using three classifiers of

TABLE V
FEATURE SETS FOR SEMG AND SOUND DATA.

Set	Feature description
Set 1	Absolute sum of 4 partitions, sum of pos/neg. deviation, Absolute sum of signal greater than 1SD, Length of absolute signal greater than 1SD, Peak count, peak distance, peak maximum
Set 2	All features from feature set 1, Spectral power, bandwidth, Centre of gravity, roll-off point, Spectral fluctuation, Sum of lin. band energy (4 bands), Log. band energy (4 bands),

different complexity: 1) Naive Bayes (NB), 2) k-Nearest Neighbour (KNN) and 3) Hidden Markov Models (HMMs). For the Naive Bayes a feature preprocessing using Linear Discriminant Analysis (LDA) was applied. The LDA filter method permitted the integration of larger feature sets and improved the discrimination performance in some situations as described below.

For the KNN classifier $k = 10$ was chosen, however only a minimal performance degradation was observed for $k = 5$. For the HMMs, continuous left-right models with 5 states were used for each class with one Gaussian mixture per feature. Continuous features were derived according to feature set 2. To reduce the influence of the varying training performance, 10 instances of each HMM were trained and tested on the training data. The best performing set was used for the evaluation.

Training and testing was performed on the subject-specific data sets. To account for variations in the partitioning of classifier training and testing data set a 10-fold cross-validation procedure was used. For training 9 of 10 parts of all instances were used. Evaluation was performed on the left out data part in order to test every instance exactly once.

The chosen nutrient groups resulted in class skews (one class contained more instances than another class). To avoid training a skewed classifier an equal number of training instances was used for all classes and the test instances were adapted to satisfy the cross-validation procedure as described before.

To compare the classification results the normalised accuracy was used:

$$\text{Normalised accuracy} = \frac{1}{2} * \left(\frac{TP}{P} + \frac{TN}{N} \right). \quad (3)$$

The normalised accuracy is robust against skew with a given (trained) classifier [32]. In our evaluation, the classifiers were trained with an equal class distribution. The measures are derived from the two-class confusion matrix as seen from one class: true positives (TP), all positives (P), true negatives (TN) and all negatives (N).

1) *Class skew analysis:* We present a performance analysis that incorporates the class skew based on class-wise precision and recall metrics. The class skew analysis simulates different class distributions using the classification result of the full evaluation dataset. The procedure starts with all relevant instances from class 1 and adds instances from the second

class sequentially. This procedure is repeated for class 2 by stepwise removing instances from class 1. For each class distribution precision and recall was computed. The results are presented in the following class skew plots. Precision and recall were derived for each class in the same way as for the event detection described in Section III.

B. Classification results

The class distribution for three volume and viscosity categories as presented in hypotheses 1 for volume and viscosity respectively (see Tab. IV), performed weak on all tested classifiers, sensor streams and feature sets. Therefore we concentrated on the evaluation of hypotheses 2 (classes for low and high volume/viscosity).

1) *Volume classification:* Fig. 5 illustrates the classification performance using the class skew precision-recall plot procedure as described before. The midpoint of each curve shows the performance for the class distribution in the evaluation data set. A natural distribution may be found to contain a large variation in swallowing volume, depending on nutrient, taste and physiology. According to the actual distribution the classifiers produces a result along the curves. Best performance is found towards the top-right corner (high precision, high recall).

The classification result of LDA+Naive Bayes using SEMG and sound (individually and by feature-level fusion) and from one KNN is shown in Fig. 5. For these results feature set 1 was utilised. Fig. 6 shows a comparison of the different classifiers using feature set 2. The best performing LDA+Naive Bayes classifier from Fig. 5 is shown for reference. Overall the LDA+Naive Bayes procedure with features from feature set 1 (time domain) performs marginally better than the KNN using feature from set 2. The HMMs did not improve the recognition rate compared to the best LDA+Naive Bayes.

From the graphs it can be seen that the sound data contributes largely to the discrimination result while the individual IH-EMG or combined IH-&SM-EMG classification performs relatively less using LDA+Naive Bayes. Best results are obtained from the feature-level fusion of IH-EMG and sound. Although more complex, feature set 2 did not improve the result. The classification performances for the bolus volume are summarised in Tab. VI using the normalised accuracy metric.

2) *Viscosity classification:* Fig. 7 illustrates the classification performance using the class skew precision-recall plot procedure. Similar to the volume analysis, the midpoint of each curve shows the performance for the ratio between low and high viscosity in the evaluation data set. A natural distribution may be found to contain more low viscosity swallows than obtained in the experiments of this investigation (see Tab. I). Using the non-skewed classifiers this would shift the result towards higher precision at a reduced recall.

The classification result of LDA+Naive Bayes using SEMG and sound (individually and by feature-level fusion) and from one KNN is shown in Fig. 7. For these classification results feature set 1 was utilised. The evaluation of the different classifiers using feature set 2 revealed a KNN using IH-EMG and sound features as best-performing classifier for

TABLE VI
PERFORMANCE SUMMARY FOR LOW VS. HIGH VOLUME BOLUS CLASSIFICATION.

Sensors	LDA+NB IH-EMG &SND		LDA+NB IH-EMG		LDA+NB SND		LDA+NB IH-&SM- EMG		KNN IH-EMG &SND		HMM IH-EMG &SND		HMM SND		KNN IH-EMG &SND		KNN SND	
	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
Feature set	1		1		1		1		1		2		2		2		2	
Class	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
Relevant	663	205	663	205	663	205	663	205	663	205	663	205	663	205	663	205	663	205
Retrieved	481	387	412	456	503	365	410	458	435	433	399	469	432	436	416	452	453	415
Recognised	438	162	355	148	445	147	359	154	395	165	373	179	374	147	384	173	404	156
FN	225	43	308	57	218	58	304	51	268	40	290	26	289	58	279	32	259	49
FP	43	225	57	308	58	218	51	304	40	268	26	290	58	289	32	279	49	259
Norm. acc.	0.73		0.63		0.69		0.65		0.70		0.72		0.64		0.71		0.69	

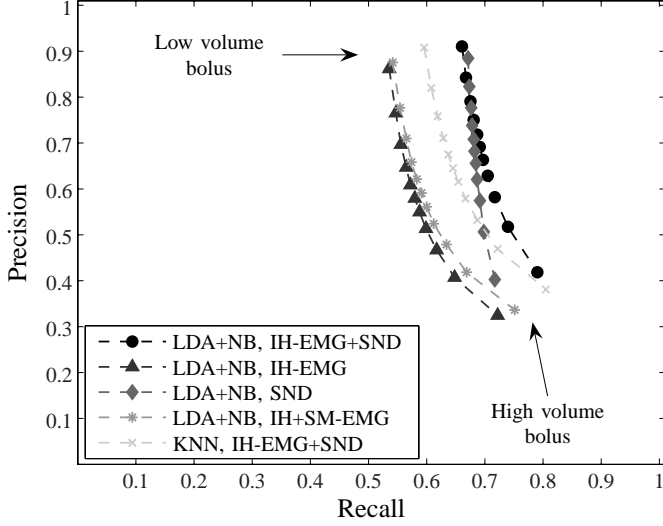


Fig. 5. Class skew plot of the low vs. high volume bolus classification result using LDA+NB and KNN. Best performance is found towards the top-right corner (high precision, high recall).

low viscosity swallows. Similar to the volume analysis, the LDA+Naive Bayes classifier performed well with feature set 1 using IH-EMG and sound. The 5-state HMMs did not improve the recognition rate compared to the best the KNN and LDA+Naive Bayes classifiers.

The classification performances for the bolus viscosity classification are summarised in Tab. VII using the normalised accuracy metric.

V. DISCUSSION AND CONCLUSIONS

The work presented in this paper aimed at 1) detecting individual swallowing events in continuous data from SEMG and sound and 2) classifying swallows regarding volume and viscosity properties.

A. Swallowing detection

For the detection of swallowing events from continuous data two approaches were presented: signal intensity thresholding and a feature similarity search. The method based on the signal intensity threshold recalled the swallowing events well (recall: 0.7), at the expense of high false positive errors (precision: 0.08). Comparably, the evaluated feature similarity

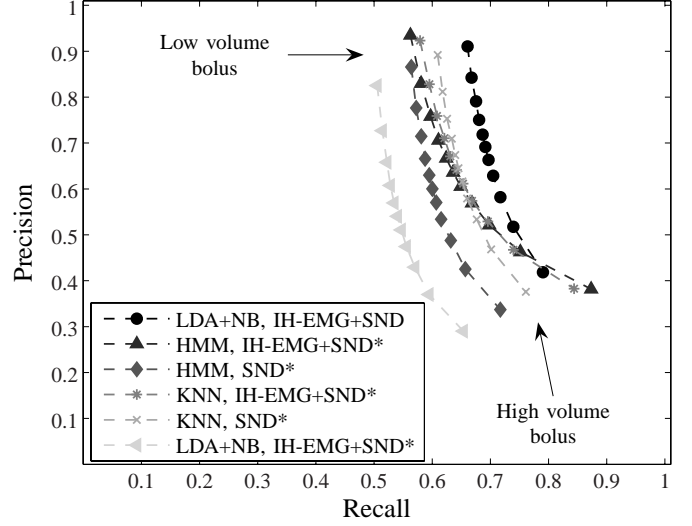


Fig. 6. Class skew plot of the low vs. high volume bolus classification result using KNN and HMM. Classifiers using feature set 2 are marked (*). The best performing LDA+NB classifier from Fig. 5 is shown for reference. Best performance is found towards the top-right corner (high precision, high recall).

methods retrieved almost half of the false positive errors, while achieving a similar recall.

The feature similarity search based on the IH-EMG signal performed better than using sound (recall and precision). However the overall result of sound is acceptable considering that the IH-EMG segmentation was used. The feature-level fusion of IH-EMG and sound similarity searches marginally improved the detection result, compared to the IH-EMG search alone.

To further improve the detection two event fusion methods were developed and tested. Both methods improved the precision clearly. However this was achieved at the expense of a reduced recall for both methods. The second-pass similarity algorithm aimed at combining the best results from the sensor-specific searches. Although a good training performance was achieved the method failed to generalise on the test data.

In conclusion, both the IH-EMG and the sound provide important information for the swallowing event detection. The feature similarity based approach to detect swallows is clearly advantageous compared to the signal intensity method.

TABLE VII
PERFORMANCE SUMMARY FOR LOW VS. HIGH VISCOSITY BOLUS CLASSIFICATION.

Sensors	LDA+NB		LDA+NB		LDA+NB		LDA+NB		KNN		HMM		HMM		KNN		KNN	
	IH-EMG &SND		IH-EMG		SND		IH-&SM-EMG		IH-EMG &SND		IH-EMG &SND		SND		IH-EMG &SND		SND	
Feature set	1		1		1		1		1		2		2		2		2	
Class	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
Relevant	558	310	558	310	558	310	558	310	558	310	558	310	558	310	558	310	558	310
Retrieved	448	420	475	393	424	444	466	402	482	386	402	466	274	594	429	439	404	464
Recognised	377	239	353	188	341	227	365	209	408	236	344	252	215	251	376	257	333	239
FN	181	71	205	122	217	83	193	101	150	74	214	58	343	59	182	53	225	71
FP	71	181	122	205	83	217	101	193	74	150	58	214	59	343	53	182	71	225
Norm. acc.	0.72		0.62		0.67		0.66		0.75		0.72		0.60		0.75		0.68	

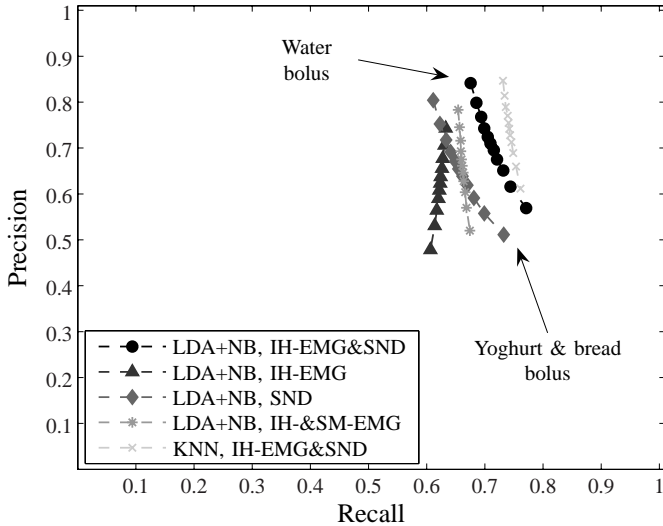


Fig. 7. Class skew plot of the low vs. high viscosity bolus classification result using LDA+NB and KNN. Best performance is found towards the top-right corner (high precision, high recall).

B. Swallowing classification

Two independent classification strategies for individual swallows were analysed: classification of bolus volume and classification of bolus viscosity. Initially the evaluated food-stuffs were grouped into three classes. However the classification result was very weak, indicating that no appropriate discriminative power was found in the sensor data and the chosen features. Therefore we concentrated on the discrimination among two classes of low and high volume as well as viscosity.

This classification revealed that the sound provides important information for volume as well as viscosity discrimination. This was expected from our initial hypothesis for the volume only. The classification result from SEMG alone was weak for both, volume and viscosity classification from the infra-hyoid and the submental positions. Best result were achieved from a feature-level fusion of IH-EMG and sound data.

We found that the combination of LDA+Naive Bayes classifier performed well given the simpler time-domain feature set. This set included static features aimed at modelling the temporal pattern of the sensor data by partitioning the complete swallow into segments. These features improved the classification result. Although without LDA, the KNN classi-

fier performed well in the evaluation. The HMMs reached the recognition rate of the best performing static classifiers.

In conclusion, a recognition rate of 0.73 to 0.75 was achieved for the volume and viscosity classifications. Although this is not an ideal performance we believe that it contributes to the envisioned dietary monitoring system. A tentative classification on individual swallowing events can be integrated since the system will be worn for entire meal consumption sessions.

VI. FURTHER WORK

From the results achieved in this work the following goals for future investigations can be derived:

- 1) Testing the methods on data from further subjects and additional nutrients to evaluate the robustness of the system and to verify the current findings regarding the classification of bolus volume and viscosity.
- 2) Evaluating the use of further sensors to improve the detection performance.
- 3) Studying the detection performance of double-swallowing and sequential swallowing specifically.

ACKNOWLEDGEMENT

This work was supported by the Swiss State Secretariat for Education and Research (SER).

REFERENCES

- [1] O. Amft, H. Junker, and G. Tröster, "Detection of eating and drinking arm gestures using inertial body-worn sensors," in *ISWC2005: IEEE Proceedings of the Ninth International Symposium on Wearable Computers.*, October 2005, pp. 160–163.
- [2] O. Amft, M. Stäger, P. Lukowicz, and G. Tröster, "Analysis of chewing sounds for dietary monitoring," in *UbiComp 2005: Proceedings of the 7th International Conference on Ubiquitous Computing.*, M. Beigl, S. Intille, and J. Rekimoto, Eds. Springer: Lecture Notes in Computer Science, September 2005, pp. 56–72.
- [3] D. M. Denk, H. Swoboda, and E. Steiner, "Physiology of the larynx," *Radiologe*, vol. 38, no. 2, pp. 63–70, Feb 1998, in German.
- [4] G. Chi-Fishman and B. C. Sonies, "Effects of systematic bolus viscosity and volume changes on hyoid movement kinematics." *Dysphagia*, vol. 17, no. 4, pp. 278–287, 2002.
- [5] J. Gates, G. G. Hartnell, and G. D. Gramigna, "Videofluoroscopy and swallowing studies for neurologic disease: a primer." *Radiographics*, vol. 26, no. 1, p. e22, 2006.
- [6] J. B. Palmer, J. C. Drennan, and M. Baba, "Evaluation and treatment of swallowing impairments." *Am Fam Physician*, vol. 61, no. 8, pp. 2453–2462, Apr 2000.
- [7] C. Ertekin and I. Aydogdu, "Neurophysiology of swallowing." *Clin Neurophysiol*, vol. 114, no. 12, pp. 2226–2244, Dec 2003.

- [8] V. Gupta, N. P. Reddy, and E. P. Canilang, "Surface EMG measurements at the throat during dry and wet swallowing." *Dysphagia*, vol. 11, no. 3, pp. 173–179, 1996.
- [9] M.-L. Huckabee, S. G. Butler, M. Barclay, and S. Jit, "Submental surface electromyographic measurement and pharyngeal pressures during normal and effortful swallowing." *Arch Phys Med Rehabil*, vol. 86, no. 11, pp. 2144–2149, Nov 2005.
- [10] W. J. Logan, J. F. Kavanagh, and A. W. Wornall, "Sonic correlates of human deglutition." *J Appl Physiol*, vol. 23, no. 2, pp. 279–284, Aug 1967.
- [11] J. A. Y. Cichero and B. E. Murdoch, "Detection of swallowing sounds: methodology revisited." *Dysphagia*, vol. 17, no. 1, pp. 40–49, 2002.
- [12] K. Takahashi, M. E. Groher, and K. Michi, "Methodology for detecting swallowing sounds." *Dysphagia*, vol. 9, no. 1, pp. 54–62, 1994.
- [13] C. Danbolt, P. Hult, L. T. Grahn, and P. Ask, "Validation and characterization of the computerized laryngeal analyzer (CLA) technique." *Dysphagia*, vol. 14, no. 4, pp. 191–195, 1999.
- [14] D. Stevens, "Ultrasound swallow." *Br Med J*, vol. 2, no. 6154, pp. 1789–1790, 1978.
- [15] T. I. Omari, N. Rommel, M. M. Szczesniak, S. Fuentealba, P. G. Dinning, G. P. Davidson, and I. J. Cook, "Assessment of intraluminal impedance for the detection of pharyngeal bolus flow during swallowing in healthy adults." *Am J Physiol Gastrointest Liver Physiol*, vol. 290, no. 1, pp. G183–G188, Jan 2006.
- [16] H. Imam, S. Shay, A. Ali, and M. Baker, "Bolus transit patterns in healthy subjects: a study using simultaneous impedance monitoring, videoesophagram, and esophageal manometry." *Am J Physiol Gastrointest Liver Physiol*, vol. 288, no. 5, pp. G1000–G1006, May 2005.
- [17] H. Firmin, S. Reilly, and A. Fourcin, "Non-invasive monitoring of reflexive swallowing," in *Speech, Hearing and Language: work in progress*. Department of Phonetics and Linguistics, University College London (UCL), 1997, vol. 10.
- [18] Y. Yamamoto, T. Nakamura, Y. Seki, K. Utsuyama, K. Akashi, and K. Jikuya, "Neck electrical impedance for measurement of swallowing," *El Eng Jp*, vol. 130, no. 4, pp. 35–44, Jan 2000.
- [19] A. Moreau-Gaudry, A. Sabil, G. Benchetrit, and A. Franco, "Use of respiratory inductance plethysmography for the detection of swallowing in the elderly." *Dysphagia*, vol. 20, no. 4, pp. 297–302, 2005.
- [20] L. Lazarek and Z. Moussavi, "Classification of normal and dysphagic swallows by acoustical means," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 12, pp. 2103–2112, Dec. 2004.
- [21] M. Pehlivan, N. Yuceyar, C. Ertekin, G. Celebi, M. Ertas, T. Kalayci, and I. Aydogdu, "An electronic device measuring the frequency of spontaneous swallowing: digital phagometer." *Dysphagia*, vol. 11, no. 4, pp. 259–264, 1996.
- [22] A. Das, N. P. Reddy, and J. Narayanan, "Hybrid fuzzy logic committee neural networks for recognition of swallow acceleration signals." *Comput Methods Programs Biomed*, vol. 64, no. 2, pp. 87–99, Feb 2001.
- [23] A. Limdi, M. McCutcheon, E. Taub, W. Whitehead, and 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, 1989.*, 9-12 Nov. 1989, pp. 1393–1394vol.5.
- [24] S. M. Sukthankar, N. P. Reddy, E. P. Canilang, L. Stephenson, and R. Thomas, "Design and development of portable biofeedback systems for use in oral dysphagia rehabilitation." *Medical engineering & physics*, vol. 16, no. 5, pp. 430–435, Sep 1994.
- [25] R. O. Dantas, M. K. Kern, B. T. Massey, W. J. Dodds, P. J. Kahrilas, J. G. Brasseur, I. J. Cook, and I. M. Lang, "Effect of swallowed bolus variables on oral and pharyngeal phases of swallowing." *Am J Physiol*, vol. 258, no. 5 Pt 1, pp. G675–G681, May 1990.
- [26] J. L. Ruark, G. H. McCullough, R. L. Peters, and C. A. Moore, "Bolus consistency and swallowing in children and adults." *Dysphagia*, vol. 17, no. 1, pp. 24–33, 2002.
- [27] R. Ding, J. A. Logemann, C. R. Larson, and A. W. Rademaker, "The effects of taste and consistency on swallow physiology in younger and older healthy individuals: a surface electromyographic study." *J Speech Lang Hear Res*, vol. 46, no. 4, pp. 977–989, Aug 2003.
- [28] J. A. Y. Cichero and B. E. Murdoch, "Acoustic signature of the normal swallow: characterization by age, gender, and bolus volume." *Ann Otol Rhinol Laryngol*, vol. 111, no. 7 Pt 1, pp. 623–632, Jul 2002.
- [29] M. Boiron, P. Rouleau, and E. H. Metman, "Exploration of pharyngeal swallowing by audiosignal recording." *Dysphagia*, vol. 12, no. 2, pp. 86–92, 1997.
- [30] E. Keogh, S. Chu, D. Hart, and M. Pazzani, "An online algorithm for segmenting time series," in *Proceedings of the IEEE International Conference on Data Mining*, 2001.
- [31] H. Junker, P. Lukowitz, and G. Tröster, "Continuous recognition of arm activities with body-worn inertial sensors," in *Proceedings of the Eighth International Symposium on Wearable Computers, ISWC, 2004*, pp. 188–189.
- [32] M. Stäger, P. Lukowicz, and G. Tröster, "Dealing with class skew in context recognition," in *IWSAWC 2006: 6th International Workshop on Smart Appliances and Wearable Computing*, Jul 2006.