

## Adaptive activity spotting based on event rates

**Citation for published version (APA):**

Amft, O. D. (2010). Adaptive activity spotting based on event rates. In *Proceedings of the IEEE International Conference on Sensors Networks, Ubiquitous, and Trustworthy Computing, 2010, Newport Beach, CA, USA 7-9 June 2010* (pp. 169-176). Institute of Electrical and Electronics Engineers. <https://doi.org/10.1109/SUTC.2010.63>

**DOI:**

[10.1109/SUTC.2010.63](https://doi.org/10.1109/SUTC.2010.63)

**Document status and date:**

Published: 01/01/2010

**Document Version:**

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.tue.nl/taverne](http://www.tue.nl/taverne)

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[openaccess@tue.nl](mailto:openaccess@tue.nl)

providing details and we will investigate your claim.

# Adaptive activity spotting based on event rates

(Invited Paper)

Oliver Amft

Signal Processing Systems, TU Eindhoven, The Netherlands

Wearable Computing Lab., ETH Zurich, Switzerland

amft@tue.nl,

<http://www.sps.ele.tue.nl/members/O.Amft>

**Abstract**—To date many activity spotting approaches are static: once the system is trained and deployed it does not change anymore. There are substantial shortcomings of this approach, specifically spotting performance is hampered when patterns or sensor noise level changes.

In this work an unsupervised sensitivity adaptation mechanism is proposed for activity event spotting based on expected activity event rates. The expected event rate for activity spotting was derived from the generalisation metric used in information retrieval. To illustrate generalisation effects and depict relations of spotting performance and event rate, different event rates were simulated and their precision-recall spotting performance analysed. Subsequently, the sensitivity adaptation concept is presented and evaluated. For this purpose two large datasets from personal healthcare applications were considered to explore benefits and limitations of this adaptation approach: recognition of drinking motions from inertial sensors and chewing strokes from sound. Results showed up to 28% spotting performance increase for event rate adapted operation, confirming performance benefits for sensitivity adaptation. The approach will be most applicable in situations, where estimated event rate statistics show low variance and long monitoring durations allow effective sensitivity adaptations.

## I. INTRODUCTION

Activity spotting can be regarded as a special area of activity recognition, which aims at detecting pattern instances in continuous sensor data. Thus, the result of an activity spotter is a set of predicted activity pattern events from a given section of sensor data. An activity pattern event in turn, has at least known event boundaries and activity duration. Given this coarse definition, activity pattern events could have arbitrary extend. Practically, the specification of what is an activity pattern event is performed during system design. Event modelling is limited by spotting algorithm and pattern model capabilities to capture relevant event pattern properties. Activity spotting principles have received broad attention in different related fields, such as spotting gestures or sounds. Nevertheless, for activity spotting in ubiquitous systems various additional sensor modalities can be equally considered for spotting. Particular examples include the spotting of car fabrication and maintenance work [1], [2], workshop tasks [3], and dietary activities [4].

Although many further applications of activity spotting exist, the realisation of spotting systems is essentially ham-

pered by a fundamental modelling problem: activity event pattern must be detectable in a continuous stream of sensor data in absence of a complete system model. For example, a gesture event may occur in principle anytime in sensor data stream, while the remaining sensor data is arbitrary. The lack of models to describe this *embedding data* (also referred to as *NULL class*) is what distinguishes activity spotting from the general class of continuous activity recognition methods, such as multi-class classification. An ideal activity spotting algorithm and model does not include any assumptions about embedding data. Thus, an ideal spotting algorithm could operate under conditions that have not been captured during design time (time of training the system). Without a completely described system it is difficult to put a decision boundary, typically a threshold, around relevant event instances to separate them from embedding data. In addition, activity spotting requires to search for events in sensor data, which is often a processing-intensive task. In turn, robustness of a spotter may be hampered when system parameters change or event instance representations change. Missing adaptive behaviour under changing patterns renders classic activity spotting algorithms error-prone under such conditions. However, maintaining a stable adaptation process is challenging, thus often the user is involved to confirm system decisions.

To perform runtime adaptations of spotting algorithms or activity event models, the previous spotting performance could be considered as a measure of success of an current parameter set. Furthermore, adapting the sensitivity of an spotting algorithm could provide a transparent control loop, where actual pattern models are left unchanged. In this work an unsupervised parameter adaptation was considered that changes the decision threshold between relevant activity event instances and embedding data. This approach could be used autonomously, hence it does not require an operator or user to provide feedback on actual performance. Activity event models were left unchanged, thus they were used as derived during supervised training.

Various metrics exist that can be considered for adaptive feedback. Typically they depict specific details of information retrieval performance [5], [6]. However, such metrics generally utilise ground truth information, which renders

them infeasible for unsupervised algorithm or model adaptations. In contrast to search and retrieval problems in related fields, e.g. for images and text (see discussion of related works below), activity pattern events have specific properties that could provide cues for estimating current performance. In this work, the expected activity event rate is considered. The activity event rate can be viewed as a common-sense property, where denoting the deviation of currently reported event rate from an conceptually expected rate provides autonomous feedback on current spotting performance. The expectation of a particular activity event rate can be a reasonable calibration measure in different activity spotting applications. Often, an averaged event rate can be estimated if the considered timeframe to sample this rate is sufficiently large. As an example, consider spotting of drinking motions in everyday activities during average days, which is relevant to prevent dehydration [7]: it is unlikely that a system user takes more than 30 sips per hour from a container. Similarly, not drinking for 3-4 hours may be unusual as well. In both cases the decision sensitivity between relevant activity event instances and embedding data could be adapted towards an expected event rate.

#### A. Paper contributions

The activity event frequency is a specific property of activity spotting applications that is considered in this work to adapt a spotting algorithm. This work focuses on evaluating benefits of such sensitivity adaptation. Thus, this approach is semi-supervised: activity event pattern models and parameters were initially derived in a supervised training step. Subsequently, an unsupervised adaptation of a spotter's sensitivity was performed by modifying the decision boundary between recognised activity events and embedding data. The following specific contributions are made:

- 1) Properties of activity spotting performance estimation and generalisation are discussed with regard to related fields of information retrieval. This is done to highlight relations of the selected spotting algorithm feedback approach with retrieval metrics. Section II addresses the fundamental properties of activity spotting. Section III illustrates the challenge of activity spotting generalisation in two datasets.
- 2) An approach for adaptive activity spotting based on expected activity event frequencies is proposed. Adaptation was performed through a control loop that modified a spotter's decision threshold. The performance of this adaptation technique was subsequently evaluated in relation to a static spotter performance. Section IV presents this adaptation approach and Section V evaluates its performance in two datasets representing complex activity pattern events from personal health monitoring.

This work presents a novel approach to obtain self-adaptive activity spotting systems that implement an unsu-

pervised control loop. Moreover, this work performs first evaluations of an event rate-based adaptation concept to estimate potential benefits of adaptive activity spotting. However, this work does not attempt to validate dynamic robustness of self-adaptive systems and applicability of this approach for any activity spotting application. The work focuses on introducing the concept of event rates and related generalisation-dependent performance analysis in the activity spotting domain.

#### B. Related work

Pattern spotting problems have been considered in various domains, most prominently in motion and gesture recognition, sign language interpretation, and sound analysis. Examples of motion and gesture recognition include immersive gaming [8], [9], and many forms of computer interaction, e.g. [10], [11], [12]. Sign language recognition is a closely related topic and considered in several works, e.g. [13]. A recent review on gesture recognition and its applications was compiled by Mitra and Acharya [14]. Systems for acoustic word spotting were considered, e.g. in [15]. In these applications a static algorithmic approach was typically considered to extract words as individual events. After an initial system training, algorithm and pattern models were not further adapted.

Semi-supervised and unsupervised techniques have been less frequently considered for pattern spotting. Recently, Yang et al. [16] proposed a threshold adaptation technique for spotting sign language based on conditional random fields (CRFs). In their approach, a threshold CRF model is constructed from state and transition feature functions of gesture CRFs. Their approach to threshold modelling is conceptually similar to the work of Lee and Kim [17], who used hidden Markov models (HMMs). Adaptation of algorithms and models after an initial training was generally not considered in these works. Yang et al. and further groups, e.g. Alon et al. [18] combat signal noise, including fractals, co-articulation, and sub-gestures using additional pattern filters. While these techniques can improve system performance, they are primarily applicable for motion and gesture recognition. In contrast, this work aims at developing an adaptive spotting technique applicable for different sensor modalities and applications.

Due to the challenges in acquiring annotated data in activity recognition, several approaches have been investigated to minimise annotation needs and thus data cost. Unlabelled samples have been considered for active learning, e.g. to ask a user to annotate key examples of activities [19]. Guan and colleagues have used co-learning and noise-removal learning to achieve sparse training datasets [20]. Stikic and Schiele have explored a similar technique, called multi-instance learning [21]. In contrast, this work focuses entirely on the performance outcome rather than modifying pattern models. However, it can be expected that active learning

techniques are complementary to the approach presented in this work.

## II. PROPERTIES OF ACTIVITY SPOTTING GENERALISATION

Information retrieval metrics, such as *precision* and *recall* are frequently considered to assess performance of activity spotters. This section summarises the relation of these metrics with regard to activity spotting. It highlights specific aspects that many current considerations of activity spotting systems lack, namely to quantify performance *generalisation*. Generalisation, in turn is tightly coupled to the activity spotting event rate, which is shown in this sections.

### A. Normalised performance metrics

Individual retrieval metrics can only partly depict spotting performance and several of these metrics are required to completely capture system performance [22], [3]. They are characterised using the following key parameters:

- 1) number of relevant items  $c$ ,
- 2) number of irrelevant items  $e$ ,
- 3) number of retrieved items  $s$ .

The number of retrieved items ( $s$ ) can be further distinguished in the number of correctly recognised items ( $s_{Correct}$ ) and the number of incorrectly returned ones ( $s - s_{Correct}$ ).

To express retrieval performance in terms of its expected success precision and recall are widely considered. Those metrics are derived as follows:

$$\text{Precision: } p = \frac{\text{number of correctly recognised items } s_{Correct}}{\text{number of retrieved items } s} . \quad (1)$$

$$\text{Recall: } r = \frac{\text{number of correctly recognised items } s_{Correct}}{\text{relevant items } c} , \text{ and} \quad (2)$$

With respect to activity spotting, these parameters can be interpreted as follows: relevant items are those activity events that have been conducted by a subject. Retrieved items represent the events that have been reported by a spotting algorithm. A correctly recognised item is a relevant activity event that has been retrieved.

The most severe errors of activity spotters are insertions and deletions, which increase the number of incorrectly returned items and lower the number of correctly recognised items respectively. Substitutions represent another severe error, however these can be neglected when activity spotters in their principal form are considered, one activity event class vs. embedding data.

### B. Activity spotting generalisation

In a general class of information retrieval systems, the total number of items is described as database size, thus  $D = (c + e)$ . Moreover, in a typical retrieval application, the number of irrelevant items broadly exceeds the number of relevant ones, thus  $e \gg c$ . Here, activity spotting requires a conceptual diversion: while activity pattern events correspond to parameter  $c$  (number of relevant items), the number of irrelevant items ( $e$ ) corresponds to embedding data, which cannot be counted in a comparable manner.

For activity spotting, embedding data is arbitrary sensor data that lacks structural information. As a consequence, different approaches have been used to deal with it. Both relevant and irrelevant items can be considered based on continuous time or timeframe results [3], [4]. Alternatively, relevant items can be viewed as countable event instances, while irrelevant and the entire database are further considered as time-continuous (in the absence of similar countable items). This latter approach is used here to describe the activity spotting generalisation. Moreover, this approach is supported by precision and recall, which both are referencing countable parameters only.

The metric *generalisation* which is used in retrieval systems to measure the embedding data size considered in an evaluation is:

$$\text{Generalisation: } g = \frac{\text{number of correctly recognised items } s_{Correct}}{\text{database size } D} . \quad (3)$$

A number of important observations with respect to generalisation and activity spotting can be made:

- While precision and recall are widely used for evaluating recognition performance of activity spotters, their performance assessment always depends on generalisation, hence the embedding data size. Thus, generalisation or embedding data size, e.g. as the amount of arbitrary data included, should always be reported in spotting evaluations. As exception, if different spotters are to be compared on the same dataset, generalisation remains constant and could be omitted.
- The generalisation metric in its isolated form is not helpful to characterise an activity spotting systems since both,  $s_{Correct}$  and  $D$  dependent on dataset and spotting algorithm. E.g. low generalisation performance is obtained for high  $D$  as well as low  $s_{Correct}$ .
- When considering the dataset size in time domain, as detailed above, generalisation is derived as a frequency: number of events per time unit. This allows to conveniently interpret generalisation as *event rate* in practical activity spotting applications.

The mutual dependency of precision and recall can be observed from Eq. 1 and Eq. 2. How generalisation influences both, precision and recall is illustrated with the help

of activity spotting datasets in Section III below.

### III. GENERALISATION IMPACT ON PERFORMANCE

The generalisation, or more conveniently, the event rate crucially influences precision-recall (PR-) performance results. This effect can be illustrated in the widely used PR-plots. In this section two activity spotting datasets and one spotting algorithm is considered to discuss spotting performance depending on generalisation.

#### A. Activity spotting datasets

1) *Dataset 1: spotting drinking gestures*: Spotting and interpreting drinking motions with a simple wearable sensor, while being exposed to other arbitrary activities virtually at the same time, is a challenging issue. In this dataset a 3-dimensional acceleration and gyroscope sensor was considered. The sensors were attached to the user’s wrist to monitor drinking motions using different containers. Scenario and analysis using a static spotting algorithm had been previously reported for this dataset in [7]. Here, the dataset properties are briefly summarised.

The dataset consists of totally 5.84 hours of sensor data from six student subjects who randomly in time took 560 sips from four different containers (glass, bottle, beer mug, cup) and four broadly scripted scenarios. The scenarios were office work, eating, gaming, and leisure. The sensor unit was sampled at 50 Hz. The recording procedure took around 50–60 min per participant. Of those, accumulated time for drinking (fetch motions) was about 11 min per subject. The remaining time contained other non-relevant activities. Overall, this dataset included 66.8 min of relevant drinking motions. Embedding data accounted for >90% of the entire dataset.

2) *Dataset 2: spotting chewing strokes*: Chewing is an important feature of the eating microstructure and relevant in various biomedical investigations and assessments, as well as to estimate food intake routines. In this dataset a ear-worn sound transducer was considered to monitor food chewed between teeth. The vibration of broken foods is conducted through mandible and skull to the ear canal. Scenario and analysis using a static spotting algorithm had been previously reported for this dataset in [23]. Here the dataset properties are briefly summarised.

Eight volunteer students (two female, six male) aged between 20 to 35 years were recruited to consume different foods in their habitual style, while chewing sounds were recorded. Sound was recorded at 44.1 kHz. The total dataset was 8.64 h. The average recording per participant was 64.83 min. In total 7910 chewing cycles were identified and annotated in 504 chewing sequences. Here we consider two example participants, who were consuming apples. Embedding data was  $\sim 97.7\%$  on average for these cases.

#### B. Considered activity spotting algorithm

Activity event spotting was performed using the feature similarity search (FSS) algorithm [4], [24]. The algorithm uses continuous sensor streams to spot events that are embedded in arbitrary data and can cope with variable-length events. This section briefly outlines FSS.

1) *Feature processing and selection*: A general set of time-domain features (for dataset 1) and frequency-domain features (for dataset 2) was computed to model event data patterns, as described in [7] and [23] respectively.

In both datasets the feature set was computed for three evenly distributed sections of sensor data, and the entire event instance. To select relevant features a Mann-Whitney-Wilcoxon test was used to compare event instances to embedding data of a training set. This ranking was refined by analysing correlations among all features. A set of 20 features (for dataset 1) and 40 features (for dataset 2) was selected that yielded the highest rank and minimum correlation scores. This approach corresponds to a method described in [25].

2) *Feature similarity search (FSS)*: The FSS algorithm consists of a signal pattern modelling (training) and a search stage. In both datasets one category of activity events was considered: drinking motions and apple chewing strokes respectively. A separate training dataset was used to determine FSS model parameters and select the feature set. An independent evaluation set was subsequently used to determine performance results.

For the FSS algorithm we used an equidistant segmentation of 0.5 Hz for dataset 1 and 0.125 Hz for dataset 2. These setting provided sufficient resolution for drinking “Sip” motions (length was  $\geq 3$  s) and chewing strokes (length was  $\leq 1$  s).

While spotting performance could be improved using data-adaptive segmentation approaches, this work emphasises the generalisation and spotter adaptation analysis. The FSS algorithm operation can be summarised as follows [4]: For each of the segmentation points, a scaling window of previously received data is analysed. The features in this window are compared to a trained model and the Euclidean distance  $d_S$  to this model was computed. A sliding window is maintained to capture temporal collisions between previously retrieved events and new ones. Each retrieved event is associated with its respective model distance  $d_S$ . A distance threshold  $\theta_d$  is used to determine sections that are considered as spotting result (retrieved items). This threshold and the scaling window search bounds were determined during the training step.

The decision step in FSS can be formulated as a function  $h_{FSS}$  to be evaluated at each timestep  $t$ :

$$t : h_{FSS} = \begin{cases} \text{report event,} & \text{if } d_S \leq \theta_d \\ \text{ignore,} & \text{otherwise} \end{cases} . \quad (4)$$

### C. Generalisation analysis

The generalisation effect was analysed by arbitrary changing dataset sizes. In particular, to demonstrate effects of varying PR-performance, the number of relevant items was changed with a constant set of embedding data.

Figure 1 shows the effect of generalisation for one participant of dataset 1 exemplary. A fraction of relevant event  $c$  was applied for each PR-curve. As this PR-plot illustrates, the spotter achieves an excellent performance when left unchanged ( $r=1.0$ ,  $p=0.96$ ). However, when the number of relevant events is decreased, performance drops as well. A similar effect could be observed when increasing embedding data size with respect to relevant items (here, drink gesture events).

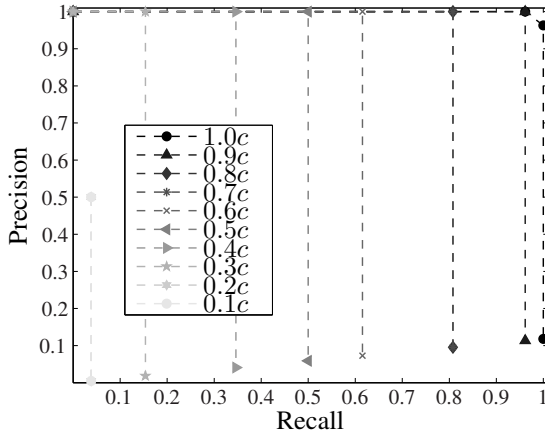


Figure 1. Drinking gesture spotting performance (dataset 1) for different generalisation settings and a constant-sized embedding data. The legend shows applied fractions of relevant events (1.0...0.1). For 1.0c the base event rate was 0.023 Hz, representing an excellent spotting performance ( $r=1.0$ ,  $p=0.96$ ).

Figure 2 illustrates an example performance for one participant of dataset 2. This plot is derived in the same way as for Fig. 1 before. However, in this case the spotter performs weaker ( $r=0.61$ ,  $p=0.55$ ), thus effects for changing the number of relevant items is not as strong as in dataset 1. At a event rate of 0.6c and below, spotting did not achieve usable results for dataset 2.

From both analyses it can be concluded that generalisation plays an important role in activity spotting analysis. On its own, generalisation can be assessed and compared with domain knowledge on the number of events per time unit to expect. In combination with PR-information, generalisation provides a more complete assessment of spotting performance. The concept of generalisation is further utilised in the following sections to adapt spotting sensitivity, given an expected event rate.

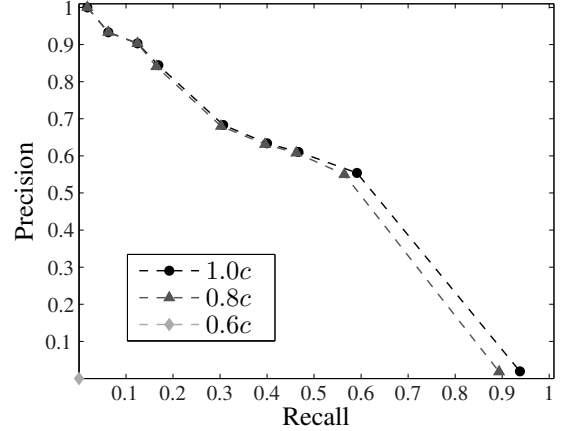


Figure 2. Chewing spotting performance (dataset 2) for different generalisation settings and a constant-sized embedding data. The legend shows applied fractions of relevant events (1.0, 0.8, 0.6). For 1.0c the base event rate was 0.069 Hz, representing a weak spotting performance ( $r=0.61$ ,  $p=0.55$ ).

## IV. CONCEPT OF EVENT RATE-ADAPTIVE SPOTTING

As shown by generalisation analysis in the previous section, spotting performance results essentially depend on selecting an appropriate relation between the number of relevant items and total dataset size. Given a constant generalisation, it is shown in this section how an FSS spotter could utilise event rate information as feedback to adapt its sensitivity.

### A. Event rate notation

Section II-B introduced generalisation for activity spotting. Specifically in activity spotting, generalisation corresponds to a easily interpretable rate of events. Given the formulation of generalisation in Eq. 3, the event rate  $E$  is subsequently measured as:

$$\text{Event rate: } E = \frac{\text{events}}{\text{time unit}} . \quad (5)$$

The separate notation is made here to distinguish the event rate as a specific measure of the activity spotting domain from a general definition of generalisation used in information retrieval. In this work, event rate  $E$  is used to measure both correctly recognised event rates (during training) and retrieved event rates (during evaluation) and denoted in Hz.

### B. Threshold-adaptive FSS

The decision boundary between relevant events and embedding data is an interpretable and influential element determining spotter performance. This is particularly relevant for FSS, which depends on estimating a decision threshold during a training step.

For FSS, a decision threshold  $\theta_d$  is derived by evaluating relevant events in a training set with respect to embedding

data. Thus the training set  $\mathcal{X}_{Train}$  must be composed of feature instances of positive (relevant events)  $\mathbf{f}_{Train}^+$  and embedding data examples  $\mathbf{f}_{Train}^-$ . Thus  $\theta_d$  is estimated by minimising the empirical training error  $\varepsilon_{emp}$ :

$$\theta_{d,Train} = \underset{\theta_d}{\operatorname{argmin}} \varepsilon_{emp}(h_{FSS}, \theta_d, \mathcal{X}_{Train}). \quad (6)$$

For the threshold-adaptive FSS,  $\theta_d$  can be dynamically modified, by minimising:

$$\theta_{d,Adaptive}(t) = \underset{\theta_d}{\operatorname{argmin}} \left( 1 - \frac{E_{Actual}(t)}{E_{Expected}} \right). \quad (7)$$

In Eq. 7, the minimisation problem is now dependent on a ratio of actual and expected (estimated from training data) event rates. Since the actual event rate is time-dependent, this holds for an adaptive threshold as well.

Eq. 7 denotes magnitude and direction of a sensitivity adaptation. The larger the ratio of  $E_{Actual}(t)/E_{Expected}$ , the larger is the estimated performance gap. Moreover, the sign in Eq. 7 denotes whether too many or too few events are reported with regard to an expected rate. Based on this concept, a standard control procedure can be deployed to perform decision threshold adaptations during system runtime.

## V. PERFORMANCE ANALYSIS OF EVENT RATE-ADAPTIVE SPOTTING

Spotter sensitivity adaptation based on the event rate criterion was subsequently analysed in two datasets, as introduced before. In particular, training-based performance is taken as baseline and performance gains achieved by the adaptation were evaluated.

In this analysis individual contributions of precision and recall are not relevant. Instead, an optimum between these two metrics was used for simplified notation and performance comparison. Thus, the f-measure was used in this section according to:

$$f = \frac{2 p r}{p + r}. \quad (8)$$

When using the decision threshold adaptation for FSS, pattern model training does not need to consider the expected event rate  $E_{Expected}$ , as long as this rate is known and can be configured during deployment. The condition of an arbitrary event rate during event model training was considered in evaluations of study participants S1-S4 of dataset 1. In participants S5 and S6 the same expected and evaluation set event rates were used. Both conditions are practically relevant, since actual event rates might fluctuate during runtime and hence require adaptations towards an expected rate.

Figure 3 shows spotting performances with regard to actual event rates  $E_{Actual}$  retrieved from the spotter. Vertical

lines in Fig. 3 indicate location of initial event rates according to the threshold training (compare Eq. 6). Subsequently, thresholds can be shifted to match  $E_{Expected}$ . Table I summarises all sensitivity adaptation performances. As these results indicate, for all participants a performance improvement or a constant performance was achieved. The largest performance increase was 28% with regard to the initial training threshold. This confirms that an expected event rate is a useful measure to estimate spotting performance.

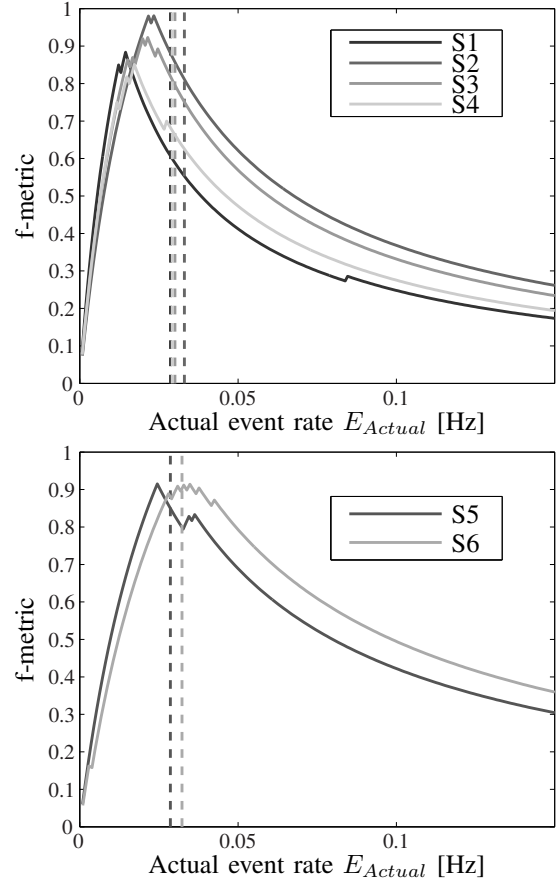


Figure 3. Drinking gesture spotting performance (dataset 1) in relation to actual event rate  $E_{Actual}$  for individual study participants (S1–S6). Performance of a training-estimated threshold is shown as vertical lines: upper plot: initially,  $E_{Expected} \neq E_{Actual}$ , lower plot:  $E_{Expected} = E_{Actual}$ .

Figure 4 shows performance results for decision threshold adaptation in dataset 2. Due to the fact that spotting performance is lower over the entire PR-space compared to dataset 1 (see Section III), adaptation gains are lower as well. This can be seen in a less pronounced maximum f-metric point. Table 4 summarises results for dataset 2. A maximum adaptation gain of 12.2% was achieved, despite the fact that initial training was performed at the expected event rate. This indicates that training and evaluation data

Table I  
SUMMARY OF THE DRINKING GESTURE SPOTTING  
PERFORMANCE (DATASET 1) FOR INDIVIDUAL STUDY  
PARTICIPANTS (S1–S6).

Metric	S1	S2	S3	S4	S5	S6
$c_{Train}$	63	66	66	74	72	76
$E_{Train}$	0.029	0.033	0.030	0.029	0.029	0.032
$c_{Test}$	23	26	25	25	34	32
$E_{Eval}$	0.017	0.023	0.02	0.02	0.029	0.033
$1 - g$ [%]	93	95	94	94	91	90
Adaptation [%] <sup>a</sup>	28	2	4	17	5	2
Peak test [%] <sup>b</sup>	34	4	5	22	12	2

<sup>a</sup> Gain in f-metric performance of event rate adaptation compared to static FSS spotting using  $\theta_{d,Train}$ .

<sup>b</sup> Difference between maximum f-metric result and static FSS spotting.

have different properties, which can be compensated by the event rate adaptation.

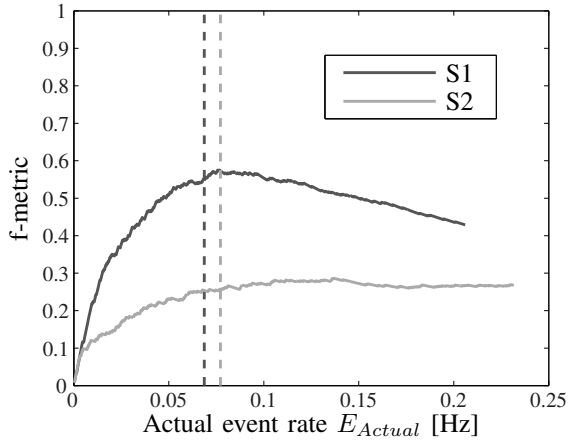


Figure 4. Chewing spotting performance (dataset 2) in relation to actual event rate  $E_{Actual}$  for individual study participants (S1, S2). Performance of a training-estimated threshold is shown as vertical lines, coinciding with  $E_{Expected}$ .

In addition, Tables II and I show differences between training-estimated thresholds and maximum f-metric performance on evaluation data. This value represents an upper bound for the adaptation gain. In practise however, as the results of this work show, this cannot be fully exploited. Obviously there are more severe pattern differences between training and evaluation, which cannot be compensated by the event rate adaptation.

## VI. CONCLUSION AND FURTHER WORK

This work presented a novel approach to adapt the sensitivity of activity spotting algorithms according to its event rate performance. Event rate is an alternate interpretation for generalisation in activity event spotting problems. Results of

Table II  
SUMMARY OF THE CHEWING SPOTTING PERFORMANCE (DATASET 2)  
FOR INDIVIDUAL STUDY PARTICIPANTS (S1, S2).

Metric	S1	S2
$c_{Train}$	225	267
$E_{Train}$	0.069	0.077
$c_{Test}$	225	267
$E_{Eval}$	0.069	0.077
$1 - g$ [%]	97.7	97.7
Adaptation [%]	12.2	-1.3
Peak test [%]	14.7	1.9

this work show that sensitivity adaptations can profoundly improve spotting performance.

A sensitivity adaptation mechanism, as it is proposed in this work, allows to obtain a transparent control loop and avoid instability. Here, adaptation was limited to event rate as single parameter for which a target (expected event rate) was specified. Moreover, contributing to system stability, an dynamically adaptive algorithm may start from a pre-trained event model configuration, which had been derived under supervised conditions.

Prerequisites for this approach are that an expected event rate can be estimated from training data, or event rate bounds can be derived based on expert knowledge of an application. It can be expected that event rate-based adaptations of activity spotters become most useful for applications where minimal event rate variance is observed. In these situations the adaptation process can compensate for variable event patterns and pattern noise levels, which hampers performance of static spotting.

This work identified generalisation as a relevant metric and useful tool for activity spotting. Performance improvements due to this self-adaptive concept are promising and suggest further exploration. In subsequent work dynamic control techniques should be considered to realise adaptive systems based on this approach.

## REFERENCES

- [1] O. Amft, C. Lombriser, T. Stiefmeier, and G. Tröster, "Recognition of user activity sequences using distributed event detection," in *EuroSSC 2007: Proceedings of the 2nd European Conference on Smart Sensing and Context*, ser. Lecture Notes in Computer Science, vol. 4793. Springer, October 2007, pp. 126–141.
- [2] T. Stiefmeier, D. Roggen, G. Troster, G. Ogris, and P. Lukowicz, "Wearable activity tracking in car manufacturing," *IEEE Pervasive Computing*, vol. 7, no. 2, pp. 42–50, April-June 2008.
- [3] J. Ward, P. Lukowicz, G. Tröster, and T. Starner, "Activity recognition of assembly tasks using body-worn microphones and accelerometers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1553–1567, Oct. 2006.



- [4] O. Amft and G. Tröster, "Recognition of dietary activity events using on-body sensors," *Artif Intell Med*, vol. 42, no. 2, pp. 121–136, February 2008.
- [5] T. Fawcett, "Roc graphs: Notes and practical considerations for researchers," HP Laboratories, Tech. Rep., 2004.
- [6] F. Provost and T. Fawcett, "Robust classification for imprecise environments," *Machine Learning Journal*, vol. 42, no. 3, pp. 203–231, March 2001.
- [7] O. Amft, D. Bannach, G. Pirkl, M. Kreil, and P. Lukowicz, "Towards wearable sensing based assessment of fluid intake," in *PerHealth 2010: Proceedings of the First IEEE PerCom Workshop on Pervasive Healthcare*. IEEE, 2010, accepted; In conjunction with the Eighth IEEE International Conference on Pervasive Computing and Communications (PerCom 2010).
- [8] D. Bannach, O. Amft, K. S. Kunze, E. A. Heinz, G. Tröster, and P. Lukowicz, "Waving real hand gestures recorded by wearable motion sensors to a virtual car and driver in a mixed-reality parking game," in *CIG 2007: Proceedings of the 2nd IEEE Symposium on Computational Intelligence and Games*, A. Blair, S.-B. Cho, and S. M. Lucas, Eds. IEEE Press, April 2007, pp. 32–39.
- [9] T. Schlömer, B. Poppinga, N. Henze, and S. Boll, "Gesture recognition with a wii controller," in *TEI 2008: Proceedings of the 2nd international conference on Tangible and embedded interaction*. New York, NY, USA: ACM, 2008, pp. 11–14, tangible and embedded interaction.
- [10] H.-K. Lee and J. H. Kim, "Gesture spotting from continuous hand motion," *Pattern Recognition Letters*, vol. 19, no. 5-6, pp. 513–520, 1998.
- [11] S. Kratz and R. Ballagas, "Gesture recognition using motion estimation on mobile phones," in *PERMID 2007: Proceedings of the 3rd International Workshop on Pervasive Mobile Interaction Devices*, May 2007, workshop at the Pervasive 2007.
- [12] S. Kallio, J. Kela, P. Korpipää, and J. Mäntyjärvi, "User independent gesture interaction for small handheld devices," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 20, no. 4, pp. 505–524, 2006.
- [13] R.-H. Liang and M. Ouhyoung, "A real-time continuous gesture recognition system for sign language," in *Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE Computer Society, April 1998, pp. 558 – 567.
- [14] S. Mitra and T. Acharya, "Gesture recognition: A survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 37, no. 3, pp. 311–324, May 2007.
- [15] Y. Takebayashi and H. Kanazawa, "Noise immunity learning for word-spotting speech recognition," *Systems and Computers in Japan*, vol. 23, no. 5, pp. 43–54, 2007.
- [16] H.-D. Yang, S. Sclaroff, and S.-W. Lee, "Sign language spotting with a threshold model based on conditional random fields," *IEEE T Pattern Anal*, vol. 31, no. 7, pp. 1264 –1277, July 2009.
- [17] H.-K. Lee and J. H. Kim, "An HMM-based threshold model approach for gesture recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 961–973, October 1999.
- [18] J. Alon, V. Athitsos, Q. Yuan, and S. Sclaroff, "A unified framework for gesture recognition and spatiotemporal gesture segmentation," *IEEE T Pattern Anal*, vol. 31, no. 9, pp. 1685–1699, 2009.
- [19] M. Stikic, K. V. Laerhoven, and B. Schiele, "Exploring semi-supervised and active learning for activity recognition," in *ISWC 2008: Proceedings of the 12th IEEE International Symposium on Wearable Computers*, 2008.
- [20] D. Guan, Y.-K. Lee, and S. Lee, "Activity recognition with the aid of unlabeled samples," in *ICUIMC 2009: Proceedings of the 3rd International Conference on Ubiquitous Information Management and Communication*. New York, NY, USA: ACM, 2009, pp. 670–674.
- [21] M. Stikic and B. Schiele, "Activity recognition from sparsely labeled data using multi-instance learning," in *LoCA 2009: Proceedings of the 4th International Symposium on Location and Context Awareness*, May 2009, pp. 156–173.
- [22] D. Huijsmans and N. Sebe, "How to complete performance graphs in content-based image retrieval: add generality and normalize scope," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 245–251, Feb. 2005.
- [23] O. Amft, M. Kusserow, and G. Tröster, "Bite weight prediction from acoustic recognition of chewing," *IEEE Trans Biomed Eng*, vol. 56, no. 6, pp. 1663–1672, June 2009.
- [24] H. Junker, O. Amft, P. Lukowicz, and G. Tröster, "Gesture spotting with body-worn inertial sensors to detect user activities," *Pattern Recogn*, vol. 41, no. 6, pp. 2010–2024, June 2008.
- [25] Q. Xu, M. Kamel, and M. M. A. Salama, "Significance test for feature subset selection on image recognition," in *International Conference on Image Analysis and Recognition*, ser. Lecture Notes in Computer Science, vol. 3211. Springer, 2004, pp. 244–252.