Systematic evaluation of social behaviour modelling with a single accelerometer
Hung, H.S.; Englebienne, G.

Published in:
UbiComp'13 adjunct: adjunct publication of the 2013 ACM Conference on Ubiquitous Computing: September 8-12, 2013, Zurich, Switzerland

DOI:
10.1145/2494091.2494130

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Systematic Evaluation of Social Behaviour Modelling with a Single Accelerometer

Hayley Hung
Technical University of Delft
The Netherlands
h.hung@tudelft.nl

Gwenn Englebienne
University of Amsterdam
The Netherlands
englebienne@gmail.com

Abstract
We describe our ongoing research on systematically analysing what types of socially related attributes and behaviours can be estimated automatically in highly social and crowded situations. This is a challenging task because obtaining the true labels for social behaviours or attributes in practice is non-trivial. Here, individuals hang a sensing device around their neck that records their acceleration during a social event. We then devise models to estimate their social behaviour or attributes based on these measurements and systematically evaluate the feasibility of such a set-up. Since we only use a single triaxial accelerometer per person, our results are surprisingly accurate and suggest that further socially relevant information could also be extracted. Our systematic evaluations provide a deeper understanding of how to better model socially relevant information in the future.

Author Keywords
Human behavior, social actions, wearable sensors, data mining

ACM Classification Keywords
H.1.2 [Models and Principles]: User/Machine Systems—Human Information Processing; H.3.1 [Information Storage and Retrieval]: Content Analysis and Indexing—Indexing Methods
Introduction

We present ongoing work [3, 4] on understanding how accurately we can automatically estimate social behaviour or attributes in crowded social settings (e.g. Figure 1 or Figure 2) using a single triaxial accelerometer, worn as a sensor badge (see Figure 3). In this context, we consider dense crowded social gatherings where their sheer size and density lead audio, video, ultra-wideband, etc. sensor data to be too noisy for robust processing.

Our premise is that the link between body motion and social behaviour has been well-documented by social psychologists [5]: so there is sufficient evidence to further study how a single accelerometer can capture this information and how effective algorithms can estimate social behaviours from this impoverished sensor setting.

Recent advances in sensor technology and signal processing have made it possible to automatically extract features reflecting body language and to relate these to the person’s state of mind, both in the lab [1, 6] and outside [7, 8]. However, up till now, these effects have been measured in relatively clean conditions and typically on a small scale [1, 9]. Large-scale experiments have been conducted [2, 8] but it relied on features, such as speech activity and proximity which still require relatively noiseless and uncrowded environments. Other related work which has estimated different aspects of social behaviour using wearable sensing devices have also relied heavily on relatively clean audio data [10, 8].

Yet interesting social behaviour also occurs in crowded and noisy situations, where identifying the speaker and robustly extracting prosodic features becomes extremely challenging. In other situations, speaking might be considered socially inappropriate (e.g. in the audience during a public presentation). Moreover, some wearers may still consider audio recordings to be an invasion of privacy, regardless of whether privacy-preserving features are used. In comparison, a single accelerometer, which could for example be worn as a conference badge, does not suffer these flaws and is easy to wear and use. Restricting the number of available sensors in each device is also appealing in terms of low battery consumption.

Practically speaking, the hardware set up we use could scale to hundreds if not thousands of users and this is the type of scenario we target.

Using this set-up, how could we still obtain valuable and useful information about the social content of an event? How can we know if the acceleration signals generated contain truly socially relevant information? Until this point, this has been considered very challenging and existing work has relied on assumed proxies for social behaviour, without systematically evaluating their performance as proxies [8, 2]. Without an understanding of the failure and success modes of such proxies, we cannot go further to improve upon the automated models of these behaviours. This extended abstract summarises a number of experiments and ongoing work to try to address this gap in our understanding so more detailed and accurate social analyses can be enabled in the future.

Data

The accelerometer data was collected using sensor badges (Figure 3(a)) hung around the volunteers’ neck using a lanyard. Each badge contains a triaxial accelerometer (recording at 20Hz onto the 4 MB of flash storage), an LCD, and are synchronised by low-power radio.

We collected data from various social events. The first event was an (inaugural) lecture where family, friends, and colleagues of the professor attended a speech (60 minutes) and drinks reception (90 minutes). About 75
hours of accelerometer data from 32 people were collected. The second event was a Symposium for 300 attendees (see Figure 2), comprising two scientific presentation sessions (80 and 100 minutes) with a short break in between, followed by a drinks reception (100 minutes). About 156 hours of accelerometer readings from 46 people were collected. Finally, a third, lab-based experiment (Mingling), (see Figure 1), was organised with 32 volunteers who were encouraged to mingle and learn about each other so that they could form quiz teams later. 10 minutes of this event was labelled for 9 people.

Experiments
We conducted experiments to evaluate our system’s ability to recognise various social attributes or behaviours:

Recognising affiliation
At the Symposium event, we gathered affiliations from the participants according to their research group. The premise was that people’s interest in a given talk should be correlated with their affiliation, which should be reflected in their body language. We used the standard k-nearest-neighbours, $k = 3$, to predict the affiliation of a person. The variance of non-overlapping 10-second windows were concatenated into a single vector for each person. Leave-one-out cross-validation resulted in a classification accuracy of 39.3%, far better than random.

Detecting Professors by how they move
We reasoned that people of varying status would have different motivations for socialising during professional social events. Therefore, professors would tend to have a larger social network and would move more between groups, while students, for example, would tend to circulate mostly within a smaller group of people.

The classification accuracies are summarised in Table 1, and were generated from the statistics of the raw magnitude (Sig) or energy (En) of the sensor readings over the drinks reception period of each event. We used a linear SVM classifier and leave-one-out cross validation with additional sub-sampling of the larger class. In the Inaugural event, the behaviour of 14 professors and 17 non-professors were available during the post talk drinks reception. For the Symposium event, data from 6 professors and 28 non-professors were recovered.

There is a notable contrast between the best two performing features. The entropy feature for the inaugural event captures both small and large motions, which represent both the gestural activities of standing people and their motion between groups. However for the mean (SigMean) feature, larger movement patterns will easily outweigh smaller movement patterns, putting more emphasis on movement between groups.

Classifying Social Actions
In the Mingling data, the actions of 9 subjects was labelled every 2s for: speaking, laughing, gesturing (either hand or head), stepping (or walking) and drinking. Using the power spectral density (PSD) for each axis of the acceleration, we trained two (multi-state) Hidden Markov Models (HMMs) for each action: one on data annotated with the ‘positive’ action and one on a random sample of sequences not associated with the ‘negative’ action. During inference, the most likely label is assigned to the observed sequence by comparing the likelihood of the data under the positive and negative HMM.

Model selection was done beforehand using different data, which was used to select the number of states per HMM, the window length over which each PSDs was calculated, and the number of frequency bins per window. The distribution parameters of the HMMs were optimised by maximum likelihood on left-out data from the same event.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Inaug.</th>
<th>Symp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SigMean</td>
<td>19.49</td>
<td>61.82</td>
</tr>
<tr>
<td>SigVar</td>
<td>41.03</td>
<td>33.22</td>
</tr>
<tr>
<td>SigKurt</td>
<td>19.11</td>
<td>35.76</td>
</tr>
<tr>
<td>SigSkew</td>
<td>22.38</td>
<td>39.35</td>
</tr>
<tr>
<td>SigEnt</td>
<td>61.27</td>
<td>35.63</td>
</tr>
</tbody>
</table>

| EnMean  | 20.65  | 59.15 |
| EnVar   | 47.83  | 34.69 |
| EnKurt  | 51.61  | 40.02 |
| EnSkew  | 42.08  | 38.74 |
| EnEnt   | **62.00** | 2.99 |

Table 1: Classification accuracy for professors and non-professors. Features were computed based on the raw acceleration signal (Sig) or its square or energy (En). Best performing features are highlighted in bold.
as the test data. We performed 10-fold cross-validation ten times on random permutations of the sequences and the classification performance is summarised in Table 2. Speaking was most detected most robustly while gesturing was the most difficult to detect. The remaining behaviours were detected with a very high precision but low recall.

We would like to exploit these behaviours to detect who is speaking with whom. Social scientists have found that people talking together have certain distinctive synchronous behaviour [5]. Preliminary analysis of our data showed that gesturing and stepping occurs synchronously more frequently, and speaking simultaneously occurs less often for people in the same group compared to different groups.

Conclusion and Future Work
We have presented systematic evaluatons of the estimation of social attributes and behaviour from accelerometer data recorded during crowded social gatherings. Future work will be focused on understanding how fusing information from the social attributes and actions could help to improve the estimation performance.

Acknowledgements
We thank researchers at the VU University of Amsterdam (Matthew Dobson, Claudio Martella, and Maarten van Steen) for the use of their wearable sensors and help during the data collection, and the University of Amsterdam (Jeroen Kools and Ben Kröse) for their data collection and annotation help.

References

Table 2: Average precision, recall and F-measure for the different action categories in our dataset over 10 repetitions of 10-fold cross validation.

<table>
<thead>
<tr>
<th></th>
<th>gesture</th>
<th>step</th>
<th>drink</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>0.59</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Recall</td>
<td>0.24</td>
<td>0.21</td>
<td>0.21</td>
</tr>
<tr>
<td>F1</td>
<td>0.34</td>
<td>0.35</td>
<td>0.35</td>
</tr>
</tbody>
</table>