



# The Three A's of Wearable and Ubiquitous Computing: Activity, Affect, and Attention

Kristof Van Laerhoven\*

University of Siegen, Siegen, Germany

**Keywords:** human activity recognition, affect recognition, attention, context awareness, ubiquitous and mobile computing

A long lasting challenge in wearable and ubiquitous computing has been to bridge the interaction gap between the users and their manifold computers. How can we as humans easily perceive and interpret contextual information? Noticing whether someone is bored, stressed, busy, or fascinated in face-to-face interactions, is still largely unsolved for computers in everyday life. The first message of this article is that much of the research of the past decades aiming to alleviate this context gap between computers and their users, has clustered into three fields. The aim is to model human users in different observable categories (alphabetically ordered): *Activity*, *Affect*, and *Attention*. A second important point to make is that the research fields aiming for machine recognition of these three A's, thus far have had only a limited amount of overlap, but are bound to *converge* in terms of methodology and from a systems perspective. A final point then concludes with the following call to action: A consequence of such a possible merger between the three A's is the need for a more consolidated way of performing solid, reproducible research studies. These fields can learn from each other's best practices, and their interaction can both lead to the creation of overarching benchmarks, as well as establish common data pipelines.

The opportunities are plenty. As early as 1960, J. C. R. Licklider regarded the symbiosis between human and machine as a flourishing field of research to come: "A multidisciplinary study group, examining future research and development problems of the Air Force, estimated that it would be 1980 before developments in artificial intelligence make it possible for machines alone to do much thinking or problem solving of military significance. That would leave, say, 5 years to develop man-computer symbiosis and 15 years to use it. The 15 may be 10 or 500, but those years should be intellectually the most creative and exciting in the history of mankind." (Licklider, 1960). Advances in Machine Learning, Deep Learning and Sensors Research have shown in the past years that computers have mastered many problem domains. Computers have improved immensely in tasks such as spotting objects from camera footage, or inferring our vital signs from miniature sensors placed on our skins. Keeping track of what the system's user is doing (Activity), how they are feeling (Affect), and what they are focusing on (Attention), has proven a much more difficult task. There is no sensor that directly can measure even one of these A's, and there are thus far no models for them to facilitate their machine recognition. This makes the three A's an ideal "holy grail" to aim for, likely for the upcoming decade. The automatic detection of a user's Activity, Affect, and Attention is on one hand more specific than the similar research field of context awareness (Schmidt et al., 1999), yet challenging and well-defined enough to spur (and require) multi-disciplinary and high-quality research. As **Figure 1** shows, the ultimate goal here is to achieve a more descriptive and accurate model of the computer's user, as sensed through wearable or ubiquitous technology.

**Activity.** The research field of wearable activity recognition has grown tremendously in the last 2 decades and can be categorized in three overlapping stages (**Figure 2**). The initial research studies focused predominantly on proving feasibility of using certain wearable sensors to automatically detect an activity, at first basic activities such as "walking" or "climbing stairs", later moving to more

## OPEN ACCESS

### Edited and reviewed by:

Kaleem Siddiqi,  
McGill University, Montreal, Canada

### \*Correspondence:

Kristof Van Laerhoven  
kvl@eti.uni-siegen.de

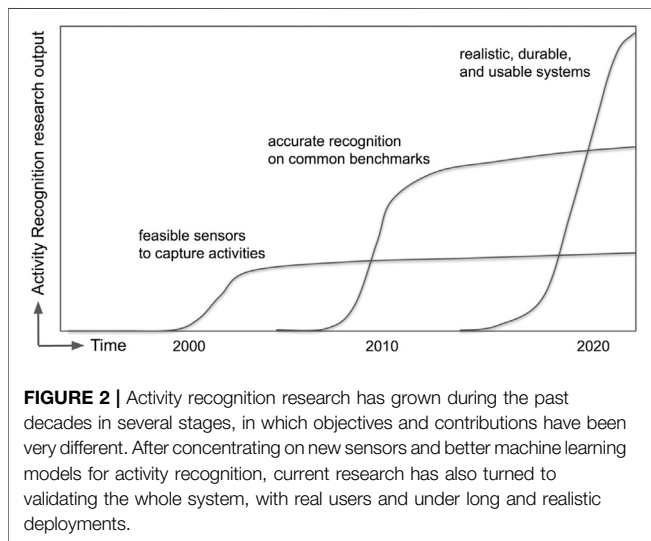
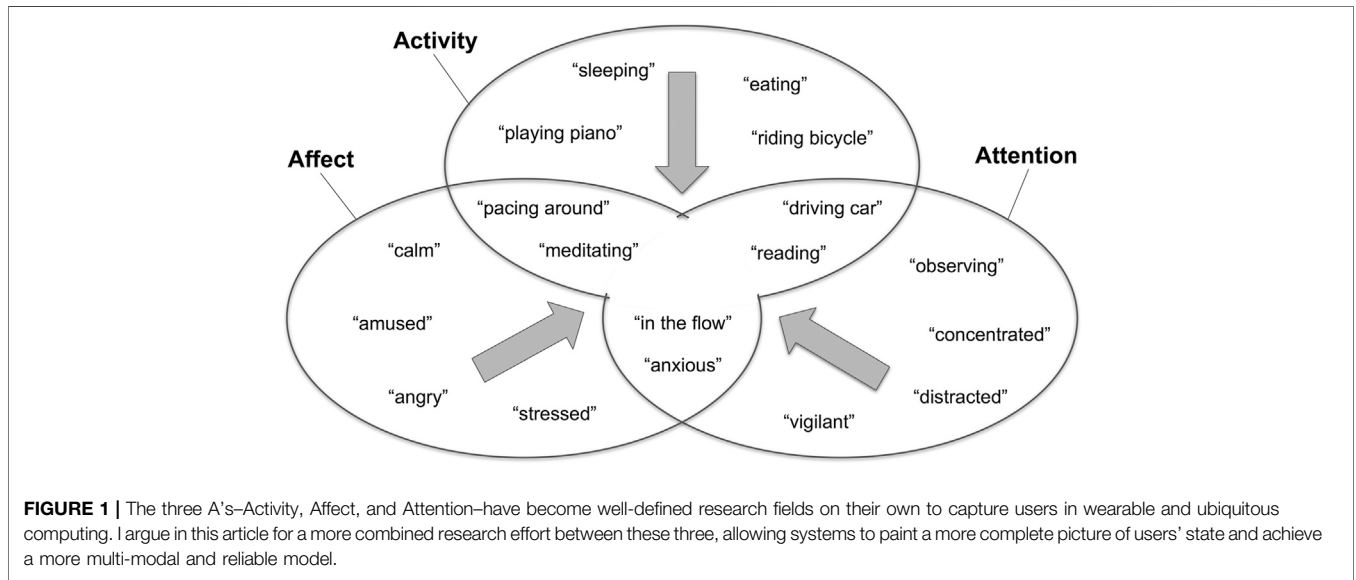
**Received:** 06 April 2021

**Accepted:** 31 May 2021

**Published:** 09 June 2021

### Citation:

Van Laerhoven K (2021) The Three A's of Wearable and Ubiquitous Computing: Activity, Affect, and Attention. *Front. Comput. Sci.* 3:691622. doi: 10.3389/fcomp.2021.691622



complex activities such as “playing tennis” or “having dinner”. This was followed up in a second step by more rigid research in data analysis and classifiers, where accuracy on benchmark datasets is key. The current, third wave of research continues with a focus on deployments with actual users, over longer stretches of time, ensuring also usability, durability, and applicability of proposed systems. Benchmark datasets in wearable activity recognition are still small, compared to other machine learning application domains, and the wide variety of types and wearing styles of predominantly inertial sensors in wearables has led to datasets that are not easy to combine. Current research in this area is going in exciting new directions to overcome these problems, however. Novel and creative ways of generating synthetic data to improve the training process for deep learning architectures include taking existing computer vision

and signal processing techniques to convert videos of human activity into virtual streams of IMU data (Kwon et al., 2020), for instance, so that much larger repositories can be used to retrieve realistic human pose sequences for given activities. Data augmentation, in which existing inertial data examples are used to create multiple different variations, possibly helped by synthetic data from physical models, is another promising direction.

**Affect.** Affect recognition has meanwhile evolved into an interdisciplinary research field bringing together researchers from natural and social sciences, and aims at detecting a person’s affective state based on sensor measurements. Applications vary from providing reasoning for a person’s decision making (for example taking camera pictures when the user has a response of being startled (Healey and Picard, 1998)), to the support of the user’s mental wellbeing (by for instance, monitoring of stress). Sensor measurements originally came mostly from microphones, cameras, or textual information, but wearable sensors have in the past years gained considerable traction in this research. Affect recognition systems that operate on imaging the user’s face can only give a temporal and spatial-limited assessment, whereas wearable-based systems can detect the user’s affective state more continuously and ubiquitously. By being able to capture increasingly more, and more reliably, both physiological and inertial data of their user, wearable systems thus represent an attractive platform for long-term affect recognition applications (Schmidt et al., 2019). The few examples of what a recognition system could detect in **Figure 1** already indicate: Affect and Activity are intertwined. Strenuous activities such as sports exercises that result in the user sweating, might wrongly cause an Affect classifier to detect a state of stress if the Activity is not considered. Likewise, affective states such as being calm or anxious might support the detection of activities like “meditating” or “pacing around”, respectively.

**Attention.** A rise of small and affordable eye tracking technologies, using fast and miniature cameras placed in the environment or closely to the user's eyes, has in the past decade given an impulse to a growing body of research in wearable attention recognition and tracking. Visual attention in the form of gaze delivers interesting information about the user, as gaze tends to signal interest—where we look also indicates what we are planning and what we have in mind. Research in human-computer interaction has also shown that gaze precedes action: Similar to how a basketball player looks at the basket before a throw, input methods have been presented in which users' gaze moving ahead of a stylus or touch input is used to quickly shift such input to where the gaze is pointing (Pfeuffer et al., 2015). A person's eye movements have been shown to be a strong sensing modality for activity recognition (Bulling et al., 2011), in particular for activities where physical activities are minimal such as reading. Modeling Attention has shown to benefit from other information such as breathing or heart rate variability, which are modalities that are also heavily used in the recognition of Affect. Bridging the user's attention model to the models of user Activity and user Affect thus not only starts out with similar sensor data, but can also mutually reinforce and improve recognition across these models.

The three A's all reflect an aspect of the user's state (physical, perceptive, cognitive, and beyond). Their relationships have been investigated and described in more depth in other research fields, particularly in Psychology, for many years. This article, however, aims at highlighting many common themes from the perspective of computer science and digital modeling, and the benefits that

such research would see in making accurate and scalable detection systems. This call to action, to combine and expand on these three fields, also includes an appeal for more common methods, data pipelines, and common benchmarks: Collecting annotated training data from realistic scenarios and large user studies is a massive undertaking, and best-practice methods to go from raw sensor signals to computational models for recognizing Activity, Affect, and Attention have been largely set up disjointly and in parallel. Annotation of sensor data is one example of a shared burden for the three A's: As these research fields have moved away from detection on well-controlled data taken under laboratory-like conditions to more realistic settings, achieving a good performance on increasingly realistic data has proven challenging. Obtaining ground truth for human activities from humans during their everyday life is as difficult as measuring states of affect, which generally is obtained from self-assessments in form of ecological momentary assessments (EMAs). Now is the time to look over the fence and explore what insights the other A's have gained and what advantages their models or joint models have to offer, in a drive to obtain a more complete and encompassing digital representation of the wearable or ubiquitous computer's user.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

## REFERENCES

- Bulling, A., Ward, J. A., Gellersen, H., and Tröster, G. (2011). Eye Movement Analysis for Activity Recognition Using Electrooculography. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 741–753. doi:10.1109/tpami.2010.86
- Healey, J., and Picard, R. W. (1998). "StartleCam: a Cybernetic Wearable Camera," in *Digest of Papers. Second International Symposium on Wearable Computers* (Pittsburgh, PA, USA: Cat. No.98EX215), 42–49. doi:10.1109/ISWC.1998.729528
- Kwon, H. H., et al. (2020). IMUTube: Automatic Extraction of Virtual On-Body Accelerometry from Video for Human Activity Recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 1–29. doi:10.1145/3411841
- Licklider, J. C. R. (1960). Man-Computer Symbiosis. *IRE Trans. Hum. Factors Electron.* HFE-1, 4–11. Available at: <http://groups.csail.mit.edu/medg/people/psz/Licklider.html>. doi:10.1109/thfe.2.1960.4503259
- Pfeuffer, K., Alexander, J., Chong, Ming, Ki., Zhang, Y., Gellersen, H., and Gellersen, Hans. (2015). Gaze-Shifting. *Proc. UIST* 15, 373–383. doi:10.1145/2807442.2807460
- Schmidt, A., Aidoo, K. A., Takaluoma, A., Tuomela, U., Laerhoven, K. V., and Velde, W. V. D. (1999). *Advanced Interaction in Context*. Heidelberg, Germany: HUC, 89–101.
- Schmidt, P., Reiss, A., Dürichen, R., Laerhoven, K. V., and Kristof, V. L. (2019). Wearable-Based Affect Recognition-A Review. *Sensors* 19 (19), 4079. doi:10.3390/s19194079

**Conflict of Interest:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Van Laerhoven. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.