Enabling Multimodal Emotionally-Aware Ecosystems Through a W3C-Aligned Generic Interaction Modality

David Ferreira[1], Nuno Almeida[1,2], Susana Brás[1,2], Sandra C. Soares[3,4], António Teixeira[1,2], and Samuel Silva[1,2(✉)]

[1] DETI - University of Aveiro, Aveiro, Portugal
sss@ua.pt
[2] IEETA – University of Aveiro, Aveiro, Portugal
[3] DEP – University of Aveiro, Aveiro, Portugal
[4] William James Center for Research, University of Aveiro, Aveiro, Portugal

Abstract. Emotions play a key role in our life experiences. In interactive systems, the user's emotional state can be relevant to provide increased levels of adaptation to the user, but can also be paramount in scenarios where such information might enable us to help users manage and express their emotions (e.g., anxiety), with a positive impact on their daily life and on how they interact with others. However, although there is a clear potential for emotionally-aware applications, they still have a long road to travel to reach the desired potential and availability. This is mostly due to the still low translational nature of the research in affective computing, and to the lack of straightforward, off-the-shelf methods for easy integration of emotion in applications without the need for developers to master the different concepts and technologies involved. In light of these challenges, we advance our previous work and propose an extended conceptual vision for supporting emotionally-aware interactive ecosystems and a fast track to ensure the desired translational nature of the research in affective computing. This vision then leads to the proposal of an improved iteration of a generic affective modality, a key resource to the accomplishment of the proposed vision, enabling off-the-shelf support for emotionally-aware applications in multimodal interactive contexts.

Keywords: Affective interaction · Generic modality · Multimodal interfaces · W3C

1 Introduction

Emotions govern our daily life experiences and much of our motivated behaviour [17,19]. The stimuli and events that capture our attention, the information we learn and recall, and the decisions we make, widely depend on the emotions we experience (e.g., happiness, sadness, fear and anxiety).

G. M. P. O'Hare et al. (Eds.): MobiHealth 2019, LNICST 320, pp. 140–152, 2020.
https://doi.org/10.1007/978-3-030-49289-2_11

The assessment of implicit measures of emotion (e.g., elevated heart rate, angry facial expression) can, nowadays, be easily performed continuously, with minimally invasive equipment and less dependent on compliance, hence offering an excellent opportunity to monitor emotional states in the context of interactive systems [5]. Gathering an insight on the user's emotional state can improve our ability to understand how the user is experiencing the environment (e.g., discomfort [1]) or task (e.g., complexity), but can also be of particular interest in aiding users understand [9], manage and communicate their emotional state, which may then result in significant improvements in their quality of life [12]. A paradigmatic example in which implicit measures of emotions may offer an added value is the case of individuals who lack their ability to communicate emotions, such as in Autism Spectrum Disorders [16]. In this line of thought, bringing the emotional state into interactive systems as an implicit interaction [14] can be a valuable resource to foster more natural and adapted forms of interaction [18]. For instance, the word "No", as an answer to a system query, is always recognised as the command "No", but, if uttered in a neutral or angry emotional state, it might be regarded differently, by the system.

Affective Computing [13] is the scientific area responsible for the study and development of systems and devices that can recognize, interpret, process, or simulate human affects. Even though the field has strongly evolved, in the past few years, and several contributions have been made in making the proposed methods available, to developers, e.g., through the proposal of affective libraries, much is yet to be done in bringing this work to its full potential in a broad set of interactive scenarios. First, developers should be able to deploy emotionally-aware interactive systems without having to master the different available technologies, or having to develop custom application logic, every time a new affective computing method needs to be integrated or a different application scenario is addressed. Second, affective computing researchers should be able to more rapidly deploy their methods into interactive scenarios, to favor the translational nature of affective computing research. This should enable a 'fast track' for the assessment of novel methods in more ecological scenarios. In this context, a solution that could be used off-the-shelf, by developers and researchers, to transparently support the design and development of emotionally-aware interactive systems would be a useful resource. To this end, and evolving a first proof-of-concept for a generic affective modality proposed in [10], this article presents the work carried out aiming at: (a) a broader vision on how affective inputs can be provided on a variety of scenarios, considering a common approach; (b) an improved conceptualisation of the modality's architecture and how it can encompass, for instance, affective fusion; (c) an enhanced compliance with the W3C recommendations for multimodal interactive systems, particularly when considering mobile scenarios; (d) an explicit consideration regarding how to deal with multiple users; and (e) a redesign of how the data required to determine the emotional state of the user reaches the modality and is subsequently processed.

The remainder of this document is organised as follows: Sect. 2 briefly presents an overview of relevant background and related work; then, Sect. 3

presents the adopted vision for the role and utility of an affective modality in the context of multimodal interactive systems and defines the main requirements to take in consideration to evolve previous work; these requirements then lead to the proposal of the enhanced modality's overall architecture presented, in Sect. 4, along with a summary of the main aspects of its development; in Sect. 5 a proof-of-concept application is presented to illustrate the integration of the modality in a simple use scenario; finally, Sect. 6 presents some conclusions and ideas for future work.

2 Background and Related Work

This section briefly provides a background on the main aspects and work deemed relevant to contextualise the presented work: affective computing, previous work, by the authors, on affective interaction deployment in multimodal ecosystems, and the support to the development of such systems.

2.1 Affective Computing

The field of affective computing studies different input variables in order to determine and understand human emotions. The extracted user's emotional state can, for instance, enable interactive systems to provide more adapted environments or react to how the user is experiencing it. Given the potential of these technologies, there is a growing interest in affective computing in the industry of video games, marketing, and mental health applications [6].

Humans communicate their emotions in many ways, and these can be inferred from a wide range of physiological and behavioural data, such as physiological signals (e.g., electrodermal activity (EDA), heart rate (HR)), speech data, text, facial expression analysis, behaviour, amongst others. With advances in the field, different technologies are now available to extract information from those humans' expressions. Although the current technologies are not yet totally confident in the results, merging information from different inputs would increase that confidence.

Different libraries are available that allow the extraction of affective information from different kinds of data. For instance Microsoft Face API[1] can extract emotion from facial images, SightCorp[2] from images or video, Tone Analyzer[3] from text, and Empath from audio. However, the libraries that are available with their pre-trained models require some effort, from developers, to include each of them in their applications, having to acquire new knowledge about each library as well.

The proposal of novel affective computing methods is a very active field of research [13] and, with the proliferation of novel sensing technologies, new methods are continuously being proposed. However, the path from the research

[1] Face API: https://azure.microsoft.com/en-us/services/cognitive-services/face/.
[2] SightCorp: https://sightcorp.com/.
[3] Tone Analyzer: https://tone-analyzer-demo.ng.bluemix.net/.

lab (e.g., with controlled data collection environments, or considering a wide range of software technologies), to real scenarios is sometimes hard to achieve. In this context, our team is not only interested in improving how developers can deploy emotionally-aware systems, but also in how translational multidisciplinary approaches to tackle complex problems, for instance regarding Mental Health, can be supported. This requires that not only affective computing researchers need to have a fast track to the interactive applications, but also that the research in sensing devices needs to be brought into the picture in a way that it can rapidly reach end-applications. How this might be articulated will be one of the key points of our conceptual vision, presented ahead.

2.2 Previous Work on an Affective Modality

In Henriques et al. [10], the authors have identified a set of challenges that need to be tackled to more effectively and profusely bring affective computing to interactive ecosystems. First, it is important to support developers in adding this feature to their applications without requiring them to master the different technologies involved; and, secondly, the research in new and improved methods for affective computing should have an easier and faster way to reach interactive systems, also with advantages on how it can be validated. This latter aspect is often hindered by a mismatch between the technologies used for developing interactive systems and those used for prototyping novel affective computing methods (e.g., Matlab). Therefore, an approach should be devised that completely decouples these two aspects and makes changing or updating the methods completely transparent to both the interactive system developers and the users.

In light of these challenges, we proposed a first affective modality, i.e., a module that could be connected to applications to provide affective inputs, and showed [10] that it was feasible to integrate the management of any available methods to extract affective information in a way that it became transparent to the developers. Nevertheless, this first effort, while extracting requirements from a broad conceptual vision of how affective interaction could be made available in complex interactive ecosystems, some requirements that would enable its easier adoption were left behind, for instance, in mobile scenarios, or the support for multiple users. Additionally, while the possibility of multiple sources of data for computing affective parameters was considered, the proposed vision did not make any consideration regarding how, conceptually, multiple affective information could be simultaneously tackled.

2.3 Developing Multimodal Interactive Systems

The development of multimodal interactive systems is a complex task, and a major effort has been done, by the W3C consortium, in proposing an architecture and a set of standards to support multimodal interaction (MMI) [7].

A few frameworks have been proposed to support the development of MMI systems, such as HephaisTK [8], OpenInterface [15], MUDRA [11], and AM4I [4]. One distinguishing characteristic of the latter is that it is based on the open

standard for MMI, which was proposed by the W3C. A simplified model of the architecture is presented in Fig. 1. The most notable aspect of the implemented architecture is that the different modules are fully decoupled: a set of modalities (i.e., ways of interacting with the system) communicate, using a standard markup, with a central Interaction Manager. The application logic receives and sends messages to and from the modalities also by communicating with the Interaction Manager [4].

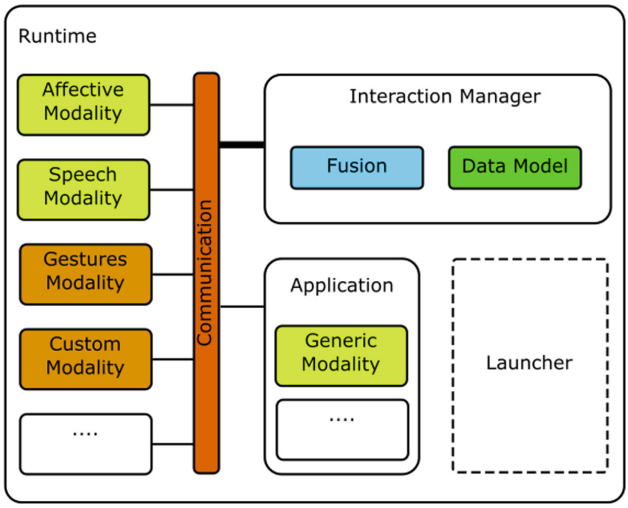

Fig. 1. Multimodal interactive architecture.

The framework reduces much of the work required to deploy interactive systems and allows the fast creation of new applications. One of its key aspects is the decoupled and modular approach: by configuring a new application with the requirements that are needed to communicate with the framework, it can immediately take advantage of its capabilities. For instance, the framework already provides an off-the-shelf modality to support speech interaction, and the developer only has to configure a few parameters with no need to master the development of speech synthesis and recognition logic. Also, the framework includes multiplatform and multi-device interaction capabilities, enabling the use of different devices, simultaneously, to interact with one application [3].

One of the interesting aspects of AM4I is that it explicitly embraces the concept of generic modalities. A generic modality, like other modalities, enables interaction with the system. Additionally, these modalities are an integral part of the framework, i.e., they come off-the-shelf when the framework is adopted and allows developers to easily deploy them with little to no effort to configure the modality. One notable example of a generic modality, which is already part of the AM4I framework, is the speech modality [2].

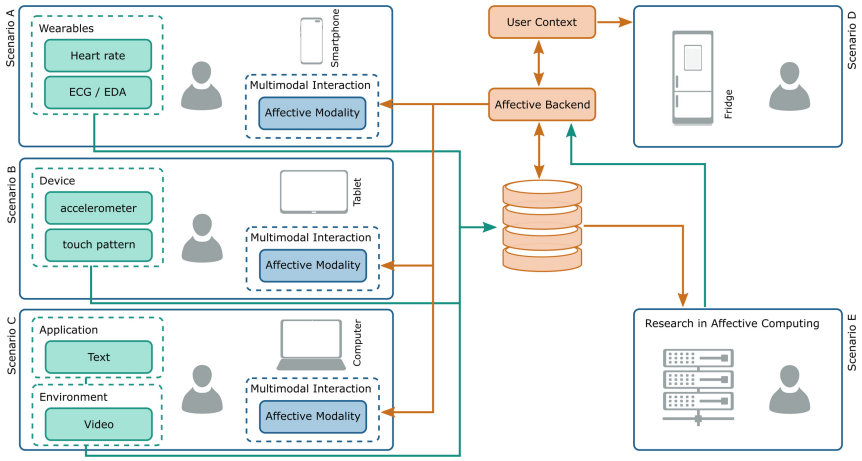

Fig. 2. Overview of the deployment of emotionally-aware applications in diverse scenarios. The sensing data and the affective computing logic are decoupled from the interactive systems, being an affective generic modality providing affective inputs to the various interactive contexts.

In this regard, any effort that results in a generic Affective Modality, which can be integrated within the framework and, thus, made available to all systems that adopt it, is an important step to bring affective interaction to a broad range of applications, for instance, in a smart-home environment.

3 Generic Affective Modality

In a first stage, we aim to have a conceptual vision on how affective computing would be deployed (and useful) in the diversity of scenarios where interaction is, nowadays, a possibility and a necessity, extending, as well, the ideas that were previously presented in [10].

3.1 Conceptual Vision on Affective Computing in a Diverse Multimodal Interactive Ecosystem

In line with what was said, Fig. 2 depicts a set of illustrative scenarios where affective interaction and affective contexts are important features.

One of the distinguishing marks of this illustrated vision is the fact that the data that is used for affective state extraction can be obtained from a wide range of sensors, which may eventually be completely decoupled from the applications. Scenarios A, B and C depict this mentioned diversity, with the sensing being performed at different levels and not only using wearables or the sensors in the user's device. For instance, a sensor in the environment, e.g., a surveillance camera, can capture facial expressions that might be used to compute the user's affective state. If the user is, for example, moving around a building, taking, at

each time, different devices with him (e.g., smartphone, tablet), the interactive ecosystem should be able to make the most out of the available sensing data to adapt itself to the obtained affective information.

Another important aspect to retain is that the interactive systems do not need to have direct access to the sensing data, but only to the extracted affective information, enabling a higher degree of data privacy. For instance, if the data used to compute the affective state comes from ECG data, no application will ever have access to it, but to the extracted affective information; or an application that is indirectly obtaining affective states, derived from locally acquired accelerometer data, will never have access to it or know where it came from.

Scenario D represents the cases where the system does not explicitly receive an affective input (i.e., from an affective modality). Instead, it can access the available affective information in the user's context, which can be populated through the features that are available in the affective backend. The main difference is that the affective information on the user's context is not necessarily instantaneous, e.g.,, it can result from a full day representative history that will enable adaptation anyway, e.g., of a smart fridge to the user's mood in that day.

Additionally, Scenario E depicts the Affective Computing research environment. Naturally, it does not entail any interaction. Instead, it provides the means for the researchers to rapidly access relevant data and deploy (improvements of) their methods into interactive ecosystems. This entails that the Affective Backend should be able to integrate novel methods in a way that is transparent for the remaining modules.

3.2 Requirements

Considering a first exploratory work [10], presenting a first proof-of-concept affective modality, and the extended context and vision presented in the previous section, we established a set of requirements that, adding to the features already considered for the first version of the modality, should be central in this novel iteration of the modality, namely:

- **Cleaner integration with mobile platforms** — The first version of the Affective Modality did not fully respect the W3C standards for Multimodal Interaction, particularly, in how the communication between the modality and the remaining modules of the architecture was performed, with some of the communication logic being tightly coupled with the application's core. Thus, to fully abide to the W3C standard, the new modality should only communicate with the Interaction Manager by resorting to MMI Lifecycle Events. Such consideration would allow the modality to be fully decoupled from the mobile systems, facilitating its integration in a wider range of scenarios;
- **Multiuser support** — To bring the Affective Modality closer to real-world scenarios, one needs to consider that multiple users may be using systems in a simultaneously way, being necessary to ensure that the modality and its backend can properly manage their different requests. This implies user identification to be possible, for cases when data is selected to identify the affective

state of a specific user and the result is made available to be consumed by its corresponding application.
- **Storage and Processing of Sensing Data** — Previously, a Firebase data storage was considered for collecting the sensing data to be subject to affective information extraction. While being a reasonable approach, for demonstration purposes, it raises several scalability and versatility concerns, considering the actual envisaged extended scenario. Therefore, a novel method should be devised to properly deal with the stream and transient natures of the arriving data and its subsequent processing;
- **Affective fusion** — Considering that multiple streams of sensing data can be simultaneously available for each user (e.g., ECG, video, text typed in an email), that means an equal number of possibilities to extract affective states. Therefore, at a particular time, multiple affective states may be available, and it is important to determine how they are considered to provide a final decision. As such, the novel iteration of the Affective Modality should be able to support affective fusion, that is, know how multiple affective states, if available, can be used together;

4 Generic Affective Modality

In light of the presented conceptual vision, we designed and implemented a generic affective modality, improving our earlier work, and adopting the features and standards of the AM4I framework [4].

4.1 Architecture

Regarding the development of the Affective Modality, three main modules (stream data processor, hub fusion, modality controller) were deemed relevant for it to properly serve its purpose, being an overview depicted in Fig. 3. Their main characteristics and roles are as follows:

- The **modality controller** is responsible for creating a session of data stream consumption whenever a client connects to it. It is responsible for beginning the stream consumption session and for managing the communication with the Interaction Manager, for instance, to convey changes of the user's affective state.
- The **stream data processor** is a module that runs in the backend and which is bound to a specific client's session, waiting for incoming streams of data that are relevant for that specific user. After a data stream is successfully consumed, it is analysed to determine its data type, e.g., GPS, video, audio, or image. With both the data type and the content in its possession, the module looks for services that can process that type of data (whether local or external) to successfully identify the user's emotional state and further relay it to the **hub fusion** module.

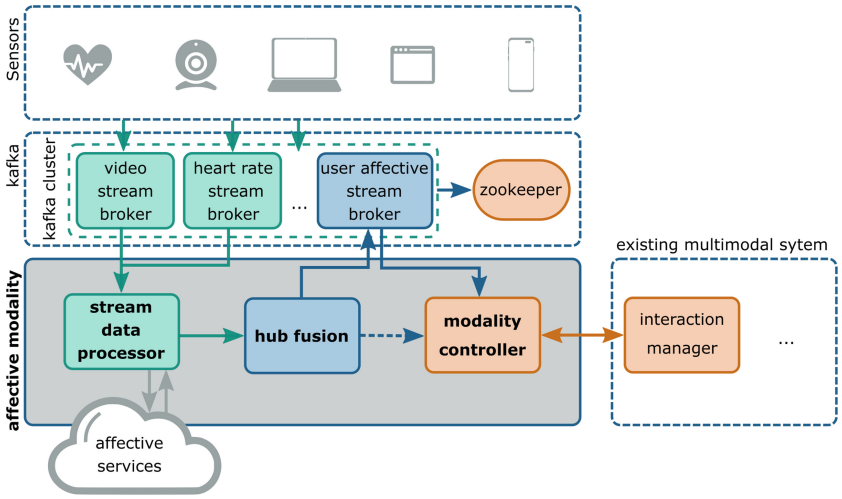

Fig. 3. Affective Modality's Architecture and supporting modules.

– The **hub fusion** module's purpose is to fuse all the events and emotions that arrive from the **stream data processor**. It tries to combine multiple affective state inputs, which result from the possibility of multiple streams of data being available for the same user, within specific time-windows, resulting in an emotion that characterises the current emotional state of the user that the inputs belong to.

4.2 Implementation

Although the mechanism to communicate the data streams from the sensors/devices is not part of the Affective Modality, it is important to note how it was implemented. For that, a distributed streaming platform, Apache Kafka[4], was considered to handle all the data streams derived from the available sensors.

These streams are subscribed by the Affective Modality, in the **stream data process**, which then identifies their types by using some supported **affective service**, such as Affectiva[5] or Kairos[6], to extract affective information. The **affective services** are a collection of computational methods, which either run locally or remotely, to support the extraction of affective information. The output of this module is delivered to the **hub fusion**, which then merges the information from different sources, and, at this stage, adopts an unsophisticated approach to choose the affective information with the highest confidence level.

After merging the available data and determining the affective state with the highest confidence level, the **hub fusion** then publishes it in a new stream, so

[4] Apache Kafka: https://kafka.apache.org/.
[5] https://www.affectiva.com/.
[6] https://www.kairos.com/.

that it can be available for consumption in other places, such as in the **modality controller** which it effectively does. This module then interprets the stream's semantics, and, if a new affective state is detected for that specific user, an MMI life cycle event, containing that data, is generated and sent to the Interaction Manager. After sending the event, all the applications that are part of the multimodal ecosystem, and which are interested in obtaining the affective state of the user to which it belongs, can consume the message so that they can adapt themselves to the detected emotional state.

4.3 Integration in the Application's Context

The versatility and decoupled nature of the used multimodal architecture enabled the Affective Modality to be easily integrated with the multimodal framework. Whenever the Affective Modality detects a change in the user's affective state, it sends a new event, containing that information, to the Interaction Manager. After the event is sent, it is the responsibility of the Interaction Manager to deliver it to any interested application, that is currently running, and in which the user that the event belongs to is currently logged.

The applications that are present in the multimodal ecosystem will continuously receive the new events that are generated by the Affective Modality, which contain the affective states of their users to allow them to adapt themselves accordingly. It is important to mention that, for the applications to be able to deploy the Affective Modality and subsequently receive their users' emotional updates, they need to implement and be able to communicate with the Interaction Manager.

5 Proof-of-Concept Results

With the integration process of the modality already described, it is now presented an illustrative proof-of-concept depicting the modality in action (Fig. 4). To this end, a simple proof-of-concept Android application has been developed, adopting the AM4I framework [4], and integrating the proposed Affective Modality. The main purpose of such integration was to enable the developed application to react to whichever user's emotional state updates received.

For the sake of simplicity, the webcam of the user's laptop is active and forwarding the video feed to his corresponding Kafka topic. The application, as soon as the user logs in, informs the modality that it has an interest in being updated regarding possible changes in its user's emotional state. When the video data reaches the user's topic, the Affective Modality backend processes it using one of the available affective services. As soon as the user's emotional state is identified, it is then redirected to the developed application so that it can react accordingly, as depicted in Fig. 4.

As seen in this practical example, by integrating the developed modality, the app developers simply need to declare that they want to know the users' affective states and deal with the semantic contents of the events that reach

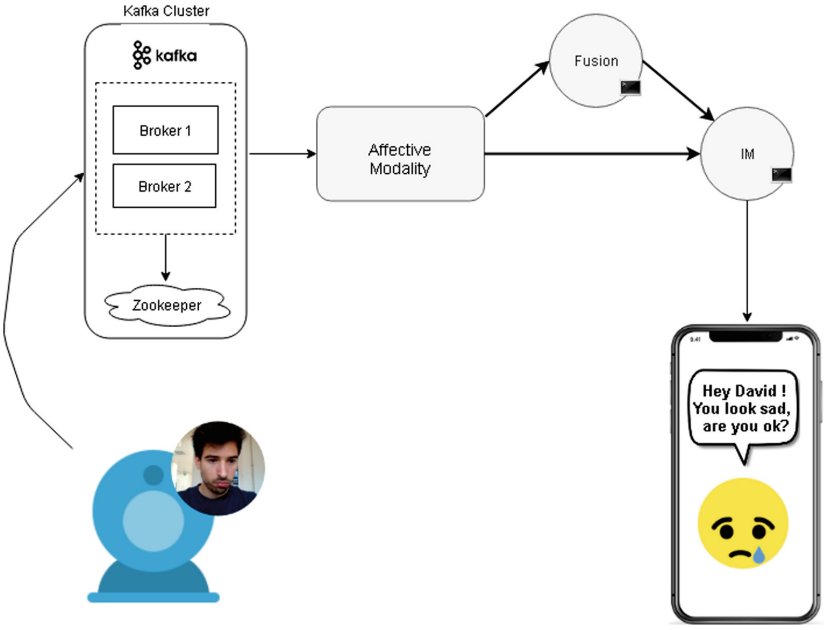

Fig. 4. Practical example of the Modality's Integration. An ambient camera captures the user's face, and that video data serves as a source for the Affective Modality to compute the associated affective state, subsequently notifying an application, on the user's smartphone, about that emotional update.

their system. This way, most of the complexity that is inherent to emotional extraction is transparently handled by remote services or toolkits, thus allowing them to focus on their system's development.

6 Conclusions

This article evolves early work, by the authors, in proposing a conceptual vision for the deployment of multimodal emotionally-aware ecosystems and showcases an improved version of a key component in this vision, a generic Affective Modality. This modality is proposed in the scope of a multimodal interactive framework, the AM4I, enabling any system that adopts it to have off-the-shelf support for affective inputs. In this regard, the full adoption of the standards proposed by the W3C also enabled a fully decoupled communication between mobile applications and the proposed modality, which was not possible in our previous proposal [10]. Additionally, the proposed vision advances how the research in affective computing can more rapidly integrate with real scenarios, and the proposal of a sensing repository based on Kafka should enable the consideration of technologies, such as Kafka Streams, to support the development of affective methods, thus facilitating their faster integration with the Affective Modality. This, should be the goal of our future efforts.

Acknowledgement. This work is partially funded by IEETA Research Unit funding (UID/CEC/00127/2019), by Portugal 2020 under the Competitiveness and Internationalisation Operational Program, the European Regional Development Fund through project SOCA – Smart Open Campus (CENTRO-01-0145-FEDER-000010), and project Smart Green Homes (POCI-01-0247-FEDER-007678), a co-promotion between Bosch Termotecnologia S.A. and the University of Aveiro.

References

1. Alavi, H.S., Verma, H., Papinutto, M., Lalanne, D.: Comfort: a coordinate of user experience in interactive built environments. In: Bernhaupt, R., Dalvi, G., Joshi, A., Balkrishan, D.K., ONeill, J., Winckler, M. (eds.) INTERACT 2017. LNCS, vol. 10515, pp. 247–257. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67687-6_16
2. Almeida, N., Silva, S., Teixeira, A.: Design and development of speech interaction: a methodology. In: Kurosu, M. (ed.) HCI 2014. LNCS, vol. 8511, pp. 370–381. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07230-2_36
3. Almeida, N., Silva, S., Teixeira, A., Vieira, D.: Multi-device applications using the multimodal architecture. In: Dahl, D.A. (ed.) Multimodal Interaction with W3C Standards, pp. 367–383. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-42816-1_17
4. Almeida, N., Teixeira, A., Silva, S., Ketsmur, M.: The am4i architecture and framework for multimodal interaction and its application to smart environments. Sensors **19**(11), 2587 (2019)
5. Barrett, L.F.: How Emotions are Made: The Secret Life of the Brain. Houghton Mifflin Harcourt, Boston (2017)
6. Brigham, T.J.: Merging technology and emotions: introduction to affective computing. Med. Reference Serv. Q. **36**(4), 399–407 (2017)
7. Dahl, D.A. (ed.): Multimodal Interaction with W3C Standards. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-42816-1
8. Dumas, B., Lalanne, D., Oviatt, S.: Multimodal interfaces: a survey of principles, models and frameworks. In: Lalanne, D., Kohlas, J. (eds.) Human Machine Interaction. LNCS, vol. 5440, pp. 3–26. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-00437-7_1
9. Gay, V., Leijdekkers, P., Agcanas, J., Wong, F., Wu, Q.: CaptureMyEmotion: helping autistic children understand their emotions using facial expression recognition and mobile technologies. Stud. Health Technol. Inform. **189**, 71–76 (2013)
10. Henriques, T., Soares, S.C., Silva, S., Almeida, N., Brás, S., Teixeira, A.: Emotionally-aware multimodal interfaces: preliminary work on a generic affective modality. DSAI 2018, 20–22 June 2018, Thessaloniki, Greece (2018)
11. Hoste, L., Dumas, B., Signer, B.: Mudra: a unified multimodal interaction framework. In: Proceedings of the 13th International Conference on Multimodal Interfaces, pp. 97–104. ICMI 2011, ACM, New York (2011). https://doi.org/10.1145/2070481.2070500
12. Picard, R.W.: Emotion research by the people, for the people. Emotion Rev. **2**(3), 250–254 (2010)
13. Poria, S., Cambria, E., Bajpai, R., Hussain, A.: A review of affective computing: from unimodal analysis to multimodal fusion. Inf. Fus. **37**, 98–125 (2017)
14. Schmidt, A.: Implicit human computer interaction through context. Personal Technol. **4**(2–3), 191–199 (2000)

15. Serrano, M., Nigay, L., Lawson, J.Y.L., Ramsay, A., Murray-Smith, R., Denef, S.: The openinterface framework: a tool for multimodal interaction. In: Proceeding of the Twenty-Sixth Annual CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI 2008, p. 3501. ACM Press, New York, April 2008. https://doi.org/10.1145/1358628.1358881
16. Sucala, M., et al.: Anxiety: there is an app for that. a systematic review of anxiety apps. Depression and Anxiety 34, 518–525 (2017)
17. Solovey, E.T., Afergan, D., Peck, E.M., Hincks, S.W., Jacob, R.J.K.: Designing implicit interfaces for physiological computing (2015)
18. Teixeira, A., et al.: Design and development of Medication Assistant: older adults centred design to go beyond simple medication reminders. Univ. Access Inf. Soc. **16**(3), 545–560 (2016). https://doi.org/10.1007/s10209-016-0487-7
19. Tyng, C.M., Amin, H.U., Saad, M., Malik, A.S.: The influences of emotion on learning and memory. Front. Psychol. **8**, 1454 (2017)