Marcos-Pablos, S., García-Peñalvo, F. J., & Vázquez-Ingelmo, A. (2021). Emotional AI in Healthcare: A Pilot Architecture Proposal to Merge Emotion Recognition Tools. In M. Alier & D. Fonseca (Eds.), Proceedings TEEM'21. Ninth International Conference on Technological Ecosystems for Enhancing Multiculturality (Barcelona, Spain, October 27th – 29th, 2021) (pp. 342-349). ACM. https://doi.org/10.1145/3486011.3486472

Emotional AI in Healthcare: a pilot architecture proposal to merge emotion recognition tools

Samuel Marcos-Pablos
GRIAL Research Group, Department of Computer Science, Research Institute for Educational Sciences. University of Salamanca
samuelmp@usal.es

Francisco José García-Peñalvo
GRIAL Research Group, Department of Computer Science, Research Institute for Educational Sciences. University of Salamanca
fgarcia@usal.es

Andrea Vázquez-Ingelmo
GRIAL Research Group, Department of Computer Science, Research Institute for Educational Sciences. University of Salamanca
andreavazquez@usal.es

ABSTRACT

The use of emotional artificial intelligence (EAI) looks promising and is continuing to improve during the last years. However, in order to effectively use EAI to help in the diagnose and treat health conditions there are still significant challenges to be tackled. Because EAI is still under development, one of the most important challenges is to integrate the technology into the health provision process. In this sense, it is important to complement EAI technologies with expert supervision, and to provide health professionals with the necessary tools to make the best of EAI without a deep knowledge of the technology. The present work aims to provide an initial architecture proposal for making use of different available technologies for emotion recognition, where their combination could enhance emotion detection. The proposed architecture is based on an evolutionary approach so to be integrated in digital health ecosystems, so new modules can be easily integrated. In addition, internal data exchange utilizes Robot Operating System (ROS) syntax, so it can also be suitable for physical agents.

CCS CONCEPTS

• **Human-centered computing** → Human computer interaction (HCI); Interaction techniques; Gestural input; Human computer interaction (HCI); Interaction techniques; Text input; • **Applied computing** → Life and medical sciences; Health care information systems;

KEYWORDS

Emotional AI, Healthcare, Digital Ecosystems, Software Architecture

1 INTRODUCTION

Transferring the latest developments in emotional artificial intelligence (EAI) to the province of healthcare is a promising approach, as it offers health care professionals with additional resources for providing support to patients and monitor their well-being.

The range of applications for EAI in healthcare is huge. In mental healthcare for example, EAI can be used to help patients understand their emotional state under stressful situations, so they can manage their emotions and handle difficult situations. Moreover, EAI may fill those information gaps when patients interact with health professionals. On the other hand, EAI can help health practitioners to increase the emotional understanding of their patients, thus being able to deliver diagnoses and treatment faster and with more accuracy. It can also help doctors to ensure patients are following a certain treatment, as well as assessing the treatment evolution.

Another application example for EAI are agents able to deliver personalized therapy. As more and more users employ technology as a faster and cheaper approach to health services, more healthcare services are being handled by computerized or robotic agents. In this sense, endowing chatbots or robotic companions with artificial empathy can have significant benefits, as it can improve the acceptance of users towards these new technologies.

Also, technology enhanced with EAI can provide better access to mental health treatments with solutions that automate talk therapy. In this way, as psychotherapy is one of the most time-consuming forms of therapy, emotional AI enhanced agents can make psychotherapy available at any time and without the need to booking an appointment in advance. Virtual therapists may also encourage patients to share their thoughts and feelings more deeply than a human therapist, as they may be perceived as safer environments for sharing personal information.

Apart from treatment, EAI can also have an important role in diagnosis. For example, being able to continuously monitor and evaluate patient emotional state can help predict their behavior, which can be critical in preventing suicide. Additionally, the data gathered from the patient can be added to health records in order to assist doctors with understanding suicide risk factors and its clinical management.

Given the growing interest in the field of emotional recognition in the area of human-computer interaction, numerous solutions have emerged in recent years aimed at making emotional recognition technologies available to inexperienced users. Although these technologies have been successfully employed in many fields such as online sales, trend analysis, or the study of user behavior in

social networks, there is still a long way to go before they are fully developed and capable enough to be used in the field of health.

This paper presents an architecture focused on the use of these emotional recognition tools for healthcare. The aim of the work presented here is to provide an architecture that allows the integration of these technologies into the digital health ecosystems previously developed by the authors in [1, 2]. In addition, and given that these emotional recognition tools are generally decoupled, in the sense that they are developed by independent companies and focus on emotional recognition in a single communication channel (e.g. voice, text, facial expression, body language, etc.), the aim of this work is to develop an architecture that allows integrating these tools to produce a combined and unified result. Moreover, to maintain the evolutionary nature of digital ecosystems, one of the main characteristics of the solution is that it permits new tools to be added when necessary, without having to introduce modifications in the architecture.

2 OVERVIEW OF CONSIDERED EMOTIONAL SOURCES

The proposed approach is based on the fundamentals of affective computing. In general terms, affective computing is computing that relates to, arises from, or influences emotions [3]. Affective computing is directly related to what is known as artificial emotional intelligence. Emotional intelligence can be defined as a set of skills that contribute to the expression of emotions in oneself, the appraisal of expressions in others, the effective regulation of emotion in self and others, and the use of feelings to motivate, plan, and achieve day-to-day actions [4].

The system focuses on the automatic (artificial) appraisal of emotions. Emotional expression appraisal goes beyond the traditional approach to emotion expression recognition from a single source, involving multi-channel analysis of emotions. Traditional single-channel emotion recognition proposals include the facial or body gesture recognition via image or video analysis, emotion recognition in speech from audio sources, or the analysis of data that comes from physiological sensors such as heart activity (EKG), brain electrical activity (EEG) or skin conductance (EDA) among others [5, 6].

The developed architecture tries to attain what is known as multimodal emotion recognition, that is, the use of multimodal data coming from different communication channels for the automatic recognition of emotions. To do so, it merges different already available single-channel emotion recognition tools to produce a unified estimation of the user's emotional state. Two main channels are initially considered in this work: facial emotion recognition and speech emotion recognition. As will be shown, however, the proposed architecture can be extended to incorporate other emotional information sources with little tunning.

2.1 Facial emotion recognition

The automatic recognition of emotions from facial expressions can be subdivided into three major tasks: facial detection, facial features extraction and facial expression interpretation.

Facial detection consists in locating faces in images. Many different methods have been proposed for automatic facial detection in the literature over the years, and current state of the art provides accuracies which are similar or even surpass human capabilities [7]. In addition, latest proposals are less and less influenced by the position of the face (preferably frontal and upright) as well as by the lighting conditions [8].

Once the face has been detected, facial features extraction aims to extract distinct facial characteristics. There exist two major approaches in this stage: geometric features extraction and appearance features extraction. The former provides the shape and location of facial features, while the latter searches for changes in facial appearance such as fiducial points position (e.g., corner of the lips) or skin texture (e.g., wrinkles, furrows). Generally, both approaches are combined as to detect modifications from a precomputed "neutral" facial state.

As for facial expression interpretation, the objective is to provide with meaning the observed facial features and changes, thus describing the psychology behind the displayed facial expression. Two main methods can be considered: sign judgment method and message judgment method. The sign judgment method looks for the codification of the performed expression without an interpretation of the emotion displayed. Different codification approaches have been proposed in the literature, but one of the most widely used is the e Facial Action Coding System (FACS) developed by Ekman and Friesen [9]. As a final goal, FACS looks to recognize and categorize action units (AUs) which represent the minimum units of muscular activity that can produce momentary changes in facial appearance. Automatic AU recognition provides as output one or more detected micro-movements of facial muscles along with the intensity (from a neutral state) of each AU.

On the other hand, message judgment methods attribute emotions to the detected facial changes. The goal is to parameterize emotions so that emotion labelling can be done using quantitative computational techniques. These methods have evolved from discrete models of emotion to multidimensional models. Simple discrete models associate facial expressions to the basic core of 6 emotions recognized universally (namely: anger, fear, surprise, disgust, joy, and sadness), whereas multidimensional approaches parameterize emotions as a lineal combination of different psychological dimensions. One of the approaches most widely used has been the circumplex model of affect, proposed by James A. Russell [10]. In the circumplex model emotions can be categorized by 2 dimensions: valence, from unpleasant (negative) to pleasant (positive); and arousal, from passive (weak emotion) to active (strong emotion). By varying the values of each dimension, emotions can be plotted on two coordinate axes. A problem with Russell's model is that some emotions such as fear and anger are located very close in his model, so other dimensions have been added to the model. One of the most common added dimensions is dominance, which ranges from submissive to dominant and is related to the user's control while displaying an emotion (see Figure 1).

Nowadays, there exist many different available tools able to automatically perform each of the abovementioned tasks [7]. However, rather than trying to combine the data provided by these tools, we have focused on available tools capable of performing the three stages altogether, and which provide as an output the interpretation of the emotion contained in the displayed facial expressions. The

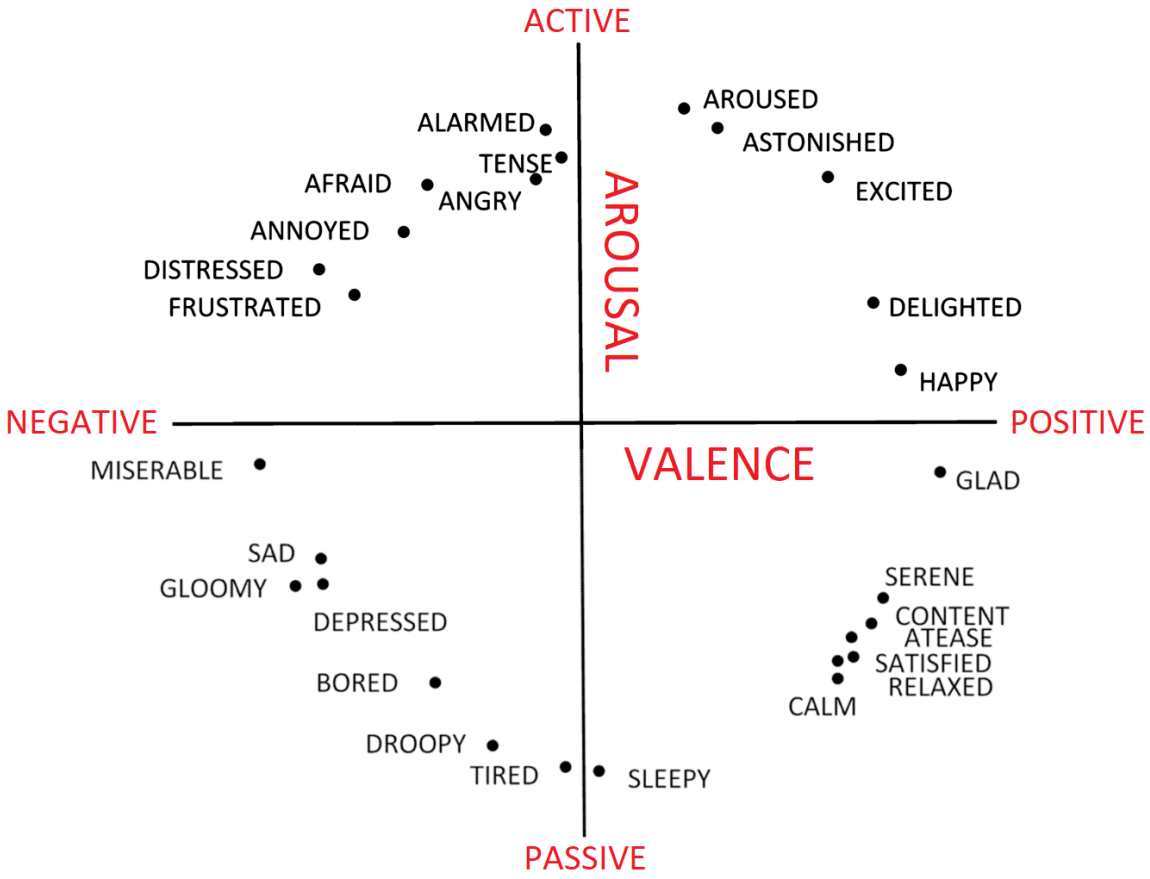

Figure 1: Russell's Circumplex model of emotions

selected cloud vision APIs have been the Affectiva Affdex SDK [11] and the Microsoft emotional API [12].

2.2 Speech emotion recognition

Recognizing emotions from speech can be approached in two complementary ways: speech signal analysis and speech content analysis. The former aims to detect emotion by analyzing the speech audio signal, trying to detect distinctive variations when the speaker is expressing a certain emotion. The latter focuses on analyzing the content of the speech (words and phrases in a particular context) looking for emotional clues in the speech entities (e.g. words,phrases).

Emotion recognition in the speech signal is carried out in two main steps: feature extraction and classification. For feature extraction, signal processing techniques are employed to produce a set of numerical values from the audio signal. These numerical values collect certain features of the audio signal so that they can be processed by a computer. The extracted features are then processed by a classifier, which is complemented by a scoring function to produce the final emotional estimation.

Extracted features aim to capture distinctive variations in speech which can be related to emotional estates. These discriminative acoustic features range from low-level descriptors to high-dimensional feature vectors. These features range from prosodic descriptors (e.g. pitch, intensity, rhythm, and duration), voice quality features (e.g. jitter), to novel features such as the nonlinear Teager–Kaiser energy operator [5]. Also, many different approaches have been proposed for classifying and scoring the extracted voice signal features, including the combination of several traditional classification algorithms such as Gaussian Mixture Model, Artificial Neural Networks, Support Vector Machines, Decision Trees or k-Nearest Neighbors. In addition, Deep Learning techniques are gaining popularity in this stage [5].

On the other hand, emotion recognition in the speech content analysis is also divided in two main stages: speech to text, where the audio signal is converted into words, and text sentiment analysis, where text is provided with emotional meaning. Speech to text process is similar to the one followed during feature extraction in emotion recognition, and it is common to perform both processes simultaneously. It includes different steps such as signal preprocessing, framing, windowing, normalization and noise reduction. After the audio signal is ready, distinctive features that correspond

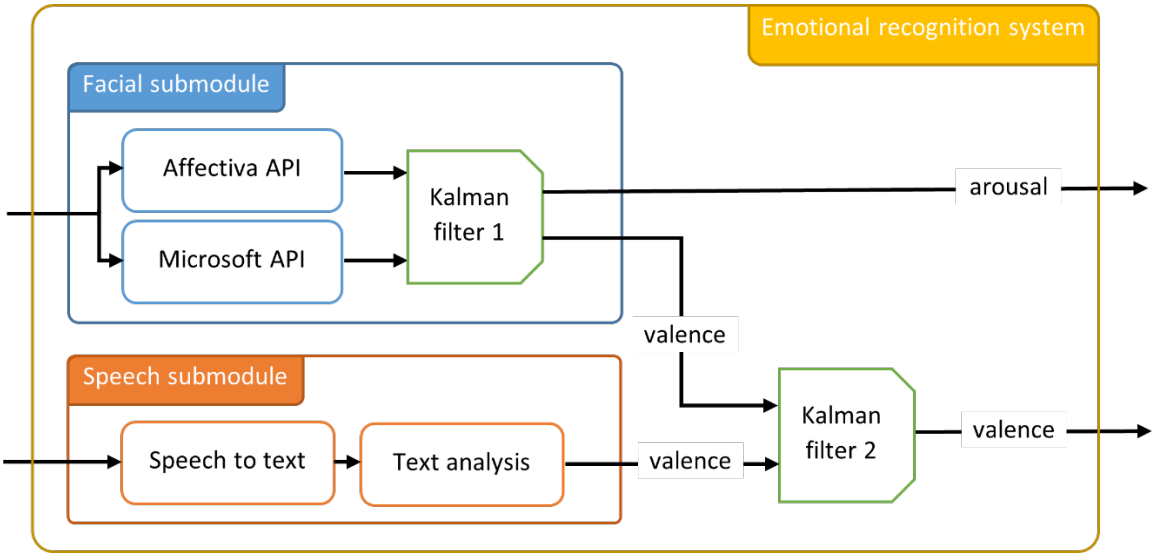

Figure 2: Emotional recognition system overview

to phonemes, words or even complete sentences are extracted, and the resulting text is presented to the user. During text sentiment analysis, natural language processing and machine learning techniques are combined to assign sentiment scores to the words, topics, themes or categories within the speech.

As happens in facial expression interpretation, speech emotion interpretation aims to provide with meaning the observed features and to describe the emotional psychology behind the speech. Among others, dimensional approaches as the one shown in Figure 2 are also employed in emotional speech recognition, which in the final term eases the process of integration with other emotional sources.

Emotion recognition in speech has been developed form many years, so many different available tools can be considered to perform this task. As in the case of facial emotion recognition, we have focused on cloud APIs which provide services that can be used on demand. The selected speech APIs have been the Bing speech API for speech to text and Microsoft Text Analytics for text sentiment analysis. It has to be noted that these tools are only focused in emotion recognition in the speech content. We have tested several other tools for emotion recognition in the speech signal such as Vokaturi, Beyond verbal or OpenSmile. However, obtained results were not sufficiently accurate to be included in this pilot architecture. As will be seen, the employed Kalman filtering approach is intended to mitigate this inconvenience, being able to extract useful information even from noisy sources. However, it requires intensive tuning to obtain proper results, so it has been leaved for future implementations.

3 PROPOSED ARCHITECTURE AND IMPLEMENTATION

3.1 Architecture overview

Figure 2 shows the proposed architecture. The adopted approach was intended to be seamlessly deployed not only in computational agents, but also in physical agents such as robots. It consists of two main submodules: the facial emotion recognition submodule and the speech emotion recognition submodule. At different stages, data coming from different sources is incrementally fused employing Kalman filters. Kalman filtering is widely used in sensor fusion integration in other fields such as robotics [8]. In fact, the developed architecture is also related to other robotic architectural implementations as data is exchanged between the different submodules employing ROS (Robot Operating System) messages [13].

The emotional recognition system modules have been programmed using both JavaScript and Python. Sound and image capture are performed through JavaScript, along with the different calls to the cloud APIs for image recognition and speech-to-emotion transcription. On the other hand, audio splitting and Kalman filtering have been deployed in Python. Python has also been employed to implement data message interchange over ROS. In addition, a front-end web page for testing purposes has been developed in html.

3.2 Facial emotion recognition submodule

As described before, the facial recognition module is composed by two components: Affectiva's Affdex SDK and Microsoft's Emotional API.

A call to the Affectiva emotion recognition service takes a video or an image as an input, and analyzes its emotional content returning a set of values using JSON syntax. After parsing the returned JSON, a value ranging from 0 to 100 for each of the six universal emotions plus contempt. This value indicates the "activation" level for each expression. Rather than using this value for emotion recognition, and in order to integrate the output with other submodules, we have employed the returned valence (which ranges from -100 to 100), and engagement (arousal) and which ranges from 0 to 100. That is, we have taken a circumplex approach to emotion recognition. The returned JSON also contains other values that

have not been considered, such as Action Units activation, an emoji representing the mood, age, gender, presence/absence of glasses and ethnicity.

On the other hand, Microsoft's emotional API takes as an image as input and returns a score for each basic emotion plus contempt. A problem that arises is how to assign values of valence and arousal to these returned values. The current approach roughly assigns valence and arousal values to each expression as described in the literature [14–17] . However more accurate approaches will be considered in future implementations [18].

3.3 Speech emotion recognition submodule

The speech emotion recognition submodule analyzes the emotions elicited by the words contained in a speech. This recognition is a two-step process composed by two sub-parts: speech to text and text analysis.

The speech to text process is performed using Microsoft Azure cloud services, particularly the Bing Speech API which has proven to be robust to background noise and little influenced by hardware variability [19]. Audio is captured continuously and when a silence in speech is detected, a chunk of raw audio data is sent to the cloud service. The cloud service responds with a JSON containing the recognized text, along with other parameters such as the detected language, recognition confidence or the duration of the speech.

This second step takes as an input the speech transcript, calls the Microsoft cognition sentiment analysis and returns a sentiment score which ranges from 0% (which represents a negative sentiment) to 100% (representing a positive sentiment). This sentiment score can be directly translated into a valence value.

Additionally, after audio is captured, it is stored as an audio file. After the generated audio file is split using the SoX - Sound eXchange audio editing software. Audio file chunks can be used for prosody analysis. Different software has been tested for prosody analysis (Vokaturi, Beyond Verbal and openSmile) but unfortunately their results were not sufficiently accurate to be integrated in the recognition system.

3.4 Message exchange

The individual components of the architecture have been developed as independent systems that communicate through the exchange of ROS Messages. In this way, different components can be activated and deactivated, or new components can be added without modifying the underlying architecture.

Developed components are defined as nodes under the ROS middleware. ROS nodes exchange data by sending and receiving messages in a publisher-subscriber fashion. The middleware provides a network on which messages are transmitted on a topic, each topic having a unique name in the ROS network. Nodes which transmit information use a publisher to send data to a topic. A node that wants to receive that information creates a subscriber to that same topic. For each defined topic, a message type is also defined, which determines the types of messages that are capable of being transmitted under that topic (e.g. floats, ints, or complex structures).

With this approach, topics allow many-to-many communication, that is, different components can send messages (publish) to the same topic and different components can receive them (subscribers). There is no need to have active subscribers for a node, data can be published and consumed at any time, and publishers and subscribers can be created and destroyed in any order.

Figure 3 provides an overview of the concept of topics, publishers, and subscribers as the messages interchanged within the facial recognition submodule. The camera capture node is external to the emotion recognition system and is responsible for capturing images from the camera and publishing them (along with the associated frame number and capture time). Both the Affectiva and Microsot nodes are subscribed to this image topic. After consuming the published message, both nodes make a call to the emotional recognition cloud services on the images, process the received JSON and then publish the results in a message that is consumed by the Kalman filtering node. This last message is an array of valence and arousal values for each of the universal expressions.

3.5 Data integration through Kalman filtering

As stated before, data integration is approached in a sensor fusion fashion, where data from different sources is combined to obtain an output which has less uncertainty than if the sources are used individually. In a sense, this approach is similar to the ones widely used in environment sensing in other fields such as robotics navigation, where map construction and robot self-location are accomplished by fusing the information from complementary sensors, redundant sensors or even from a single sensor over a period of time [8].

The Kalman filter differs from other filtering techniques in that it considers all measurements (noisy as they may be) as sources able to providing information that improves the resulting value. It is an algorithm that estimates a variable from measured data by following two steps: firstly, predicting the state of the system and secondly, incorporating the collected observations once they have been corrected. Thus, the objective is to obtain an optimal estimator, in terms of the Mean Squared Error (MSE), based on the dynamics of the system and the noisy observations. Ultimately, the Kalman filter processes all available measurements, regardless of their accuracy. The aim is to estimate the value of the variables of interest, based on knowledge of the system and observations, together with the description of their noise and errors. One of its major advantages is that it is a recursive process. This allows to incorporate new observations without having to reformulate the entire algorithm.

The mathematical implementation of the filter is beyond the scope of this paper. Briefly, since the Kalman filter requires some initial knowledge of the system (e.g. the covariance matrix of the error of the process P_k), we have estimated the parameterization of the errors of the returned values of recognition gathered from the different APIs considering the results found in the literature [8, 19, 20]. Based on this initial parameterization, the Kalman filter recursively updates the measurements obtained by the API calls with the predictions made, obtaining a filtered output. The process is shown in Figure 4

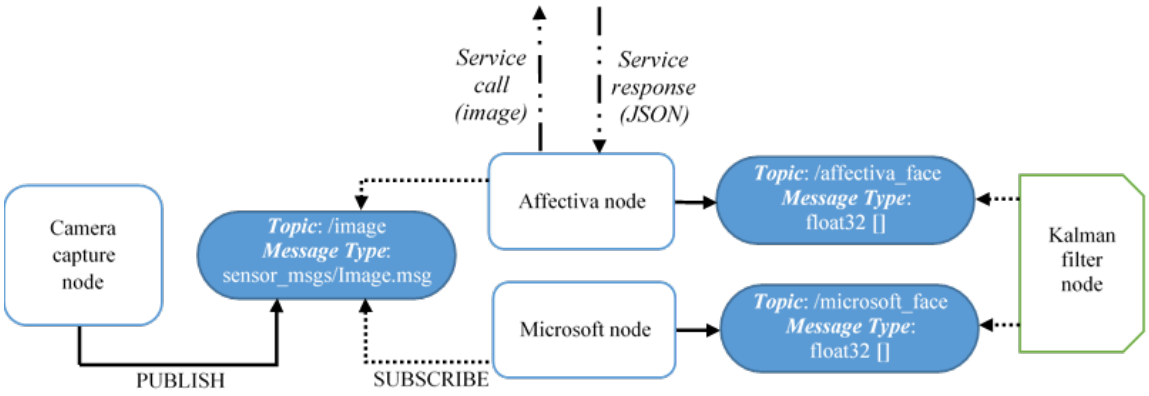

Figure 3: Exchanged messages within the facial recognition submodule.

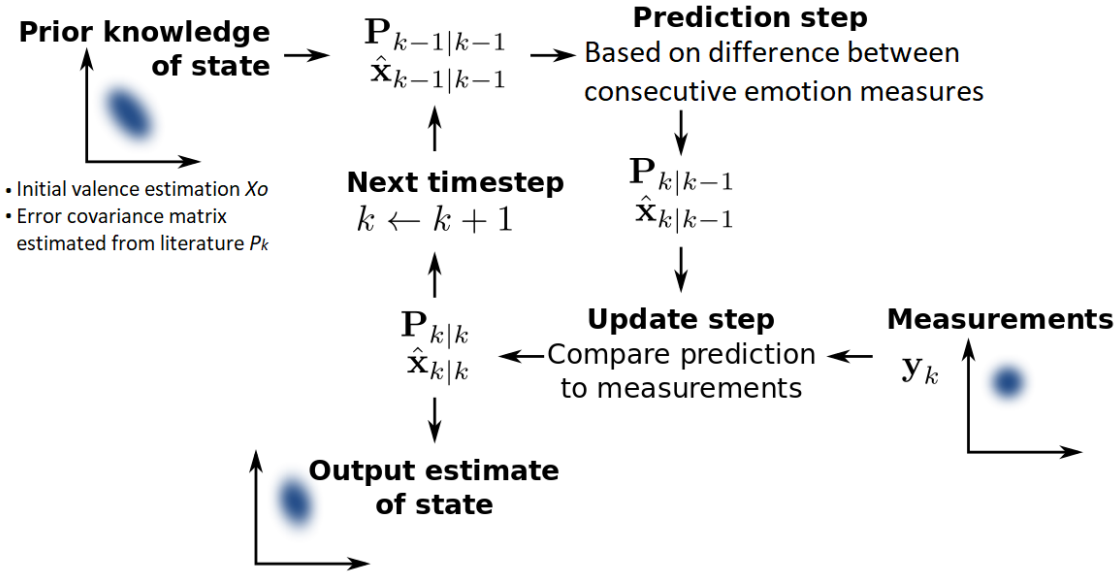

Figure 4: Kalman filtering overview

3.6 Front end

Figure 5 shows the html developed interface for testing the system. It is divided into different areas, each one allowing to control the different components (e.g. enable/disable ROS message publishing, enable/disable a certain node, etc.). The interface also includes the option to visualize real-time graphs from the results parsed from the different APIs, as well as to save all data (final and intermediate) for offline processing.

4 CONCLUSIONS AND FUTURE WORK

The future of emotional artificial intelligence looks promising and is continuing to improve. However, there are still significant challenges to tackle if AI is to be effectively used to understand and treat health conditions. One of the most important challenges is to integrate the technology into the health provision process. Because emotion AI capabilities are still under development, it is important to complement EAI technologies with expert supervision, in order to provide accurate health provision mechanisms. In this sense, the objective should be to provide health professionals with the necessary tools to make the best of EAI without a deep knowledge of the technology.

The present work aims to provide an initial architecture proposal for making use of different available technologies for emotion recognition, where their combination could enhance emotion detection. The proposed architecture is based on an evolutionary approach so to be integrated in digital health ecosystems, so new modules can be easily integrated. In addition, internal data exchange utilizes Robot Operating System (ROS) syntax, so it can also be suitable for physical agents.

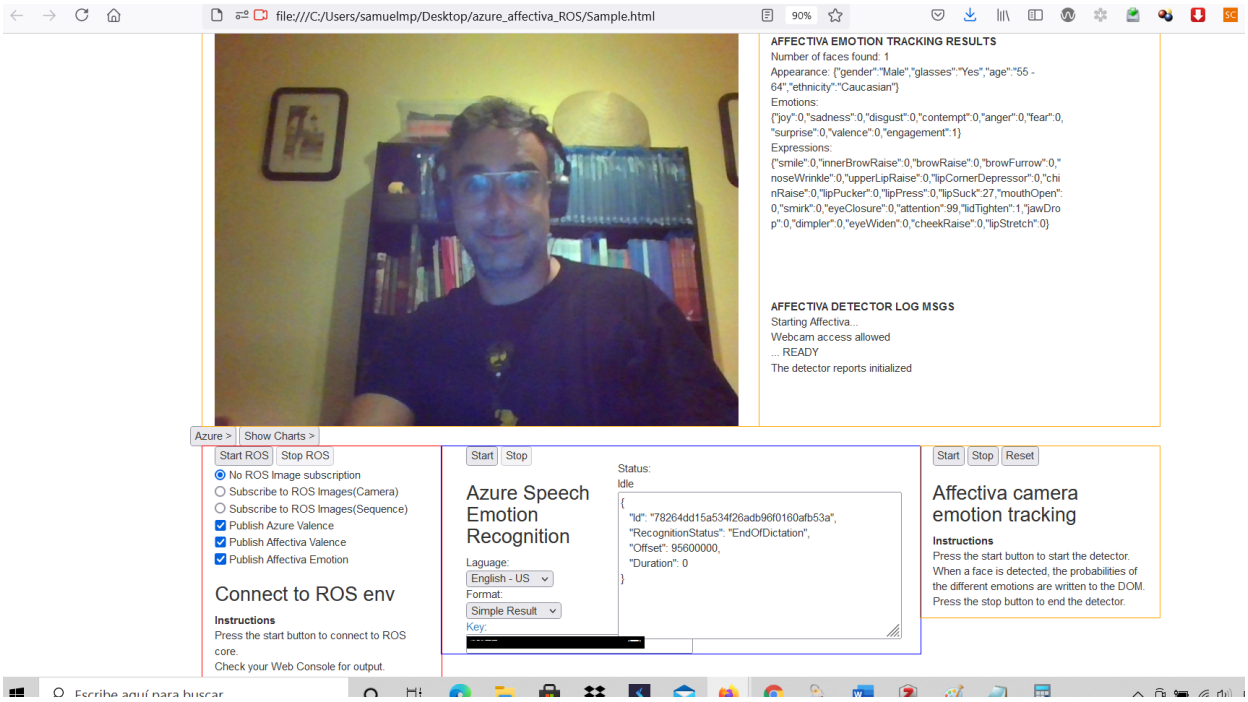

Figure 5: Developed interface for testing the proposed architecture

Future work will focus on adding new tools and emotional source channels to the proposed architecture and to fully integrate the emotion recognition architecture with our developed health ecosystem and services.

ACKNOWLEDGMENTS

This research was partially funded by the Spanish Government Ministry of Economy and Competitiveness through the DEFINES project grant number (TIN2016-80172-R) and the Ministry of Science and Innovation through the AVisSA project grant number (PID2020-118345RB-I00)

REFERENCES

[1] García-Holgado, A., Marcos-Pablos, S., & García-Peñalvo, F. J. (2019). A Model to Define an eHealth Technological Ecosystem for Caregivers. In Á. Rocha, H. Adeli, L. P. Reis, & S. Costanzo (Eds.), New Knowledge in Information Systems and Technologies (pp. 422–432). Springer International Publishing. https://doi.org/10.1007/978-3-030-16187-3_41

[2] Marcos-Pablos, S., García-Holgado, A., & García-Peñalvo, F. J. (2019). Modelling the business structure of a digital health ecosystem. Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality, 838–846. https://doi.org/10.1145/3362789.3362949

[3] Picard, R. W. (1995). Affective Computing (Technical Report No. 321). M.I.T Media Laboratory Perceptual Computing Section.

[4] Salovey, P., & Mayer, J. D. (1990). Emotional Intelligence. Imagination, Cognition and Personality, 9(3), 185–211. https://doi.org/10.2190/DUGG-P24E-52WK-6CDG

[5] Lieskovská, E., Jakubec, M., Jarina, R., & Chmulík, M. (2021). A Review on Speech Emotion Recognition Using Deep Learning and Attention Mechanism. Electronics, 10(10), 1163. https://doi.org/10.3390/electronics10101163

[6] Dupré, D., Krumhuber, E. G., Küster, D., & McKeown, G. J. (2020). A performance comparison of eight commercially available automatic classifiers for facial affect recognition. PLOS ONE, 15(4), e0231968. https://doi.org/10.1371/journal.pone.0231968

[7] Deshmukh, R. S., & Jagtap, V. (2017). A survey: Software API and database for emotion recognition. 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), 284–289. https://doi.org/10.1109/ICCONS.2017.8250727

[8] Bhattacharjee, A., Pias, T., Ahmad, M., & Rahman, A. (2018). On the Performance Analysis of APIs Recognizing Emotions from Video Images of Facial Expressions. 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 223–230. https://doi.org/10.1109/ICMLA.2018.00040

[9] Ekman, P., & Friesen, W. (2002). Facial action coding system: A technique for the measurement of facial movement. San Francisco, CA: Consulting Psychologists Press.

[10] Russell, J. A. (1980). A circumplex model of affect. Journal of Personality and Social Psychology, 39(6), 1161–1178. https://doi.org/10.1037/h0077714

[11] McDuff, D., Mahmoud, A., Mavadati, M., Amr, M., Turcot, J., & Kaliouby, R. el. (2016). AFFDEX SDK: A Cross-Platform Real-Time Multi-Face Expression Recognition Toolkit. Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, 3723–3726. https://doi.org/10.1145/2851581.2890247

[12] Del Sole, A. (2018). Introducing Microsoft Cognitive Services. In A. Del Sole (Ed.), Microsoft Computer Vision APIs Distilled: Getting Started with Cognitive Services (pp. 1–4). Apress. https://doi.org/10.1007/978-1-4842-3342-9_1

[13] Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T. B., & Leibs, J. (2009). ROS: an open-source Robot Operating System. Proc. ICRA Open-Source Softw. Workshop. International Conference on Robotics and Automation (ICRA).

[14] Hepach, R., Kliemann, D., Grüneisen, S., Heekeren, H. R., & Dziobek, I. (2011). Conceptualizing Emotions Along the Dimensions of Valence, Arousal, and Communicative Frequency – Implications for Social-Cognitive Tests and Training Tools. Frontiers in Psychology, 2, 266. https://doi.org/10.3389/fpsyg.2011.00266

[15] Paltoglou, G., & Thelwall, M. (2013). Seeing Stars of Valence and Arousal in Blog Posts. IEEE Transactions on Affective Computing, 4(01), 116–123. https://doi.org/10.1109/T-AFFC.2012.36

[16] Olszanowski, M., Pochwatko, G., Kuklinski, K., Scibor-Rylski, M., Lewinski, P., & Ohme, R. K. (2015). Warsaw set of emotional facial expression pictures: A validation study of facial display photographs. Frontiers in Psychology, 5. https://doi.org/10.3389/fpsyg.2014.01516

[17] Petrantonakis, P. C., & Hadjileontiadis, L. J. (2010). Emotion Recognition from Brain Signals Using Hybrid Adaptive Filtering and Higher Order Crossings Analysis. IEEE Transactions on Affective Computing, 1(2), 81–97. https://doi.org/10.1109/T-AFFC.2010.7

[18] Kowalczuk, Z., & Czubenko, M. (2016). Computational Approaches to Modeling Artificial Emotion – An Overview of the Proposed Solutions. Frontiers in Robotics

and AI, 3. https://doi.org/10.3389/frobt.2016.00021

[19] Mathur, A., Isopoussu, A., Kawsar, F., Smith, R., Lane, N. D., & Berthouze, N. (2018). On Robustness of Cloud Speech APIs: An Early Characterization. Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, 1409–1413. https://doi.org/10.1145/3267305.3267505

[20] Khanal, S. R., Barroso, J., Lopes, N., Sampaio, J., & Filipe, V. (2018). Performance analysis of Microsoft's and Google's Emotion Recognition API using pose-invariant faces. Proceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-Exclusion, 172–178. https://doi.org/10.1145/3218585.3224223