IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 14, 2019, accepted February 25, 2019, date of publication March 7, 2019, date of current version April 9, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2903436

Application of an Ontology-Based Platform for Developing Affective Interaction Systems

NESTOR GARAY-VITORIA, IDOIA CEARRETA, AND EDURNE LARRAZA-MENDILUZE
Informatika Fakultatea, University of the Basque Country (UPV/EHU), E-20018 Donostia, Spain

Corresponding author: Nestor Garay-Vitoria (nestor.garay@ehu.eus)

This work was supported in part by the Ministry of Economy and Competitiveness of the Spanish Government and by the European Regional Development Fund under Project TIN2017-85409-P, and in part by the Department of Education, Universities, and Research of the Basque Government under Grant IT980-16.

ABSTRACT Computer systems need to have sufficient ability and intelligence to communicate with people. To this end, they have to be able to interpret or to manage certain types of information that people are used to perceiving in human communications, such as speech modulation, facial expression, and so on taking human emotions into account. The ontology-based platform proposed in this paper attempts to support the development of resources that need to take emotion transmission into account, especially in communication between users and interactive systems. To this end, the factors relevant to the transmission of affective states have been studied and included in an ontology. Based on this ontology, a platform was created to guide the development of emotional resources that provide users with more natural interfaces. Finally, an interactive multimodal system was created to validate the proposed ontology-based platform and to apply the study to real-life cases.

INDEX TERMS Affective computing, affective recognition and synthesis, interaction context modeling, ontology knowledge representation.

I. INTRODUCTION

Multimodal interaction emerges from the need to provide users with the multiple modes of interaction required to cover their personal needs. Unfortunately, communication with many devices that people use nowadays is mainly performed by means of verbal communication (written text) and in a neutral or unemotional way. Moreover, the non-verbal information that is implicitly transmitted is often left out. This information is essential in human communication and is used for expressing our emotions. The inclusion of emotions improves the interaction by increasing people's level of understanding and decreasing the ambiguity of the messages, for example, by including emoticons. According to Mehrabian [1], about 90% of the information transmitted in human communication is non-verbal, with the verbal information transmitted accounting for only approximately 10% of the volume of information exchanged between people. Moreover, according to Picard [2], it is appreciated that these characteristics that are associated with interpersonal relationships also appear in communication with computers.

The associate editor coordinating the review of this manuscript and approving it for publication was Maurizio Tucci.

For that reason, human-computer interaction systems should be able to interpret the information coming from people and to generate a response based on this information. This has led to the emergence of the field of Affective Computing [2]–[4], which researches into the detection of and response to the user's emotions using computer-based technology. This technology could complete the development of intelligent systems by enabling them to automatically interact with users and to make their own decisions when emitting responses, without any human intervention. Thus, the common goals of users and intelligent systems can be achieved more effectively.

Currently, one of the most widely used mechanisms for modeling the knowledge of a specific domain is the use of ontologies. The main objective of ontologies is to represent real-world concepts and also the relationships between those concepts. For this purpose, it is essential to reach a consensus and to specify a common vocabulary for sharing this information. It is thus possible to share the knowledge between people (e.g. designers or developers) and between software agents (e.g. intelligent agents), and to provide these people or software agents with concepts and terms relevant to a specific domain. Moreover, ontologies have sufficient

2169-3536 © 2019 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

mechanisms for the reuse of this domain knowledge by developers who need it, without having to create a new domain [5], and they allow inferences to be made on the model instances, producing assumptions that are not easily obtained by other means. The authors have thus used ontology technology for creating a platform for guiding the development of emotional resources, taking into account the context surrounding the user. In this regard, it is essential to study and analyze the emotional and cognitive models of human beings, in order to understand their behavior and improve the systems by adapting interactions to people's personal needs and characteristics, continuing the works presented in [6]–[8].

In the following section, some related works and affect-related models are presented. These models are those that have been used in the ontology-based platform. Next, this platform is described in detail. Afterwards, a multimodal interaction system, the development of which has been based on the proposed platform, is presented and evaluated in order to prove the validity of the platform. Finally, some conclusions are drawn and future works presented.

II. RELATED WORK AND MODELS

Computing that relates to, arises from, or deliberately influences emotion or other affective phenomena is the formal description of Affective Computing, originally defined by Picard [2]. The basic idea is that interactive communication with computer systems can be significantly improved by taking the affective characteristics of human beings into account. According to Picard, the main objective is to endow computers with emotional intelligence; i.e. the ability to recognize, interpret and generate emotions.

But, *why do computers need to be able to recognize, interpret and generate emotions?* There are several fields that benefit from Affective Computing systems, including: e-learning, telemedicine, robotics and psychotherapy. For instance, in the area of e-learning, the systems can determine the emotional needs of students and can thus motivate users to learn and can keep their attention by using emotions.

Another essential question is *how can computers recognize, interpret and generate emotions?* Peter Lang proposes a model that includes three systems or communication modalities. According to [9], these communication modalities are involved in the expression of emotions and could also serve as indicators for detecting a user's emotions:

- *Verbal information:* contains the explicit message perceived or transmitted by users.
- *Behavioral:* facial and postural expressions, speech paralinguistic parameters, etc.
- *Psychophysiological responses:* such as heart rate, galvanic skin response –GSR–, and electroencephalographic response.

For instance, with regard to the speech communication modality, appropriate parameters (e.g. volume, pitch, and speed) have to be taken into account to generate or recognize emotions. This is in order to be able to either emulate diverse moods reflecting the user's affective states or, in the case of a recognizer, to create patterns for classifying the emotions transmitted by the user.

It is therefore essential to represent all this knowledge and, especially, model the emotions. The emotion theories proposed by cognitive psychology are a useful starting point for modeling affective states. The most commonly used emotion classification theories in the Human-Computer Interaction (HCI) field are the categorical [10], dimensional [11] and appraisal [12] theories. For practical reasons, categorical models of emotions have been more frequently used in Affective Computing. For example, Oudeyer [13] developed some algorithms for the production and recognition of five emotions based on speech parameters. A dimensional approach to emotion has been also advocated by a number of theorists, such as Tellegen [14]. Emotion dimensions are a simplified description of the basic characteristics of affective states [15]. The most frequently encountered emotion dimensions are Valence, Arousal and Dominance [16]. The Valence dimension is related to feeling good or bad or even giving *positive* or *negative* labels [17]. The Arousal dimension measures how excited or calm a person is. Finally, the Dominance dimension measures whether a user is in control of the situation or whether he/she is being controlled by the situation. The appraisal theoretical model offers a descriptive framework for emotion, based on the way the person experiences events, things or people at the focus of the emotional state [12].

A tool called SAM (Self-Assessment-Manikin) [16] can be used to represent or indicate an emotion based on the dimensional theory. SAM is a non-verbal pictorial assessment technique that consists of three scales corresponding to three dimensions: Valence, Arousal and Dominance. Each scale is composed of five figures that represent a human being. These scales have a range of 9 values, numbered from 1 to 9. The Valence scale of SAM describes how pleasant or unpleasant the emotion is, from left to right. The Arousal scale ranges from a state of total activity to a state of calm. In the Dominance scale, the figure on the far left represents a person who feels self-conscious, while the figure on the far right is the one that most transmits a sense of control.

Since the cognitive processes have a notable influence on affect, the researchers must also take into account which ones are involved, how they work and how they influence human-computer interaction. Some authors, including Wickens [18], consider that a human being has a cognitive system containing several sensory systems. Interaction between a human and the computer happens when there is an exchange of information. The computer presents its information in a physical way and the person must pick this up through his/her sensory systems [19]. These sensory systems are able to extract information from the environment. Perceptual processes analyze the information received through the sensory systems and assign meaning to the physical stimulus picked up by the sensory systems. Next, the information perceived is stored in the memory with the possibility of being retrieved and used later. In this case, the user generates a response using

the information retrieved from his/her memory, which is analyzed, compared and interpreted. This response is received by the computer's peripherals through its communication channels.

We also have to mention the efforts of the World Wide Web Consortium (W3C) Multimodal Interaction Working Group made in order to develop open standards for extending the Web to allow multiple modes of interaction, anyone, anywhere, any device, and any time [20]. They gave several W3C recommendations related to Affective Computing and multimodality, such as the Emotion Markup Language (EmotionML) 1.0 [21] and the Extensible MultiModal Annotation (EMMA) markup language 1.0 [22]. Based on those W3C standards, several developments have been made in order to get natural user interfaces [23].

Apart from the emotional theories, communication modalities and so on, other external factors also have to be considered in order to describe a situation that has given rise to an affective interaction between a person and an interactive system. The stimulus transmitted by a person can be analyzed (e.g. the physiological signals) in order to detect the emotion produced. However, in many cases, other aspects relating to the surrounding context could be of interest in order to provide a better understanding of this same situation.

In this regard, Göker and Myrhaug [24] propose a model where a user context is defined. In this model, apart from the personal aspects related to the user, other types of issues or elements are considered in order to describe the context of the user. This context is composed of five elements:

- *Environment context:* includes environmental data, such as topics related to the place where the user is (objects, services, temperature, light, noise, weather, etc.).
- *Personal context:* includes personal, physiological (pulse, blood pressure, weight, etc.) and mental (mood, expertise, stress, etc.) data.
- *Task context:* describes what the individuals are doing in this User context (explicit goals, actions, activities, etc.).
- *Social context:* describes the social aspects of the current User context (role with respect to friends, enemies, neighbors, etc.).
- *Spatio-temporal context:* describes aspects of the User context relating to time and space (time, location, direction, speed, etc.).

This model and most of the models explained in this section have been used for the description of the proposed ontology and the platform based on it, which are detailed in next section.

As can be seen, there are several models that represent human emotions taking one or more context elements into account [25]. Moreover, systems that recognize and generate emotions through different modalities can also be found [26]–[30]. Few of them are systems or platforms that gather the knowledge for more than one modality and various elements of the context and that also use this knowledge to provide support to the generation of affective resources, which is the idea of the proposed platform.

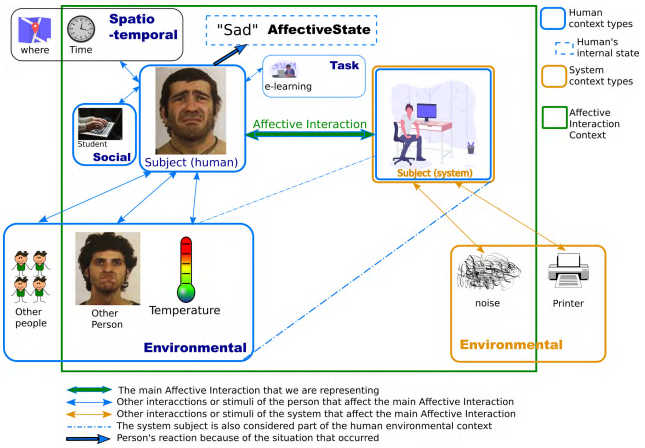

FIGURE 1. A scenario representing a context-aware affective interaction between a person and a system.

III. THE ONTOLOGY-BASED PLATFORM FOR DEVELOPING AFFECTIVE INTERACTION SYSTEMS

The main goal of the ontology-based platform proposed in this paper is to support the developments of resources that need to take emotion transmission into account in the interactions between users and systems.

Let us introduce a scenario to assist with the description of an affective interaction between a user and a system (see Figure 1). One of the goals of the proposed work is to be able to represent these types of scenarios and situations with the Affinto ontology. It is thus possible to provide a knowledge base for developing an affective system that is able to adapt itself to the user's situation.

The scenario presented in Figure 1 shows a person during a learning session in an e-learning system. This system has been developed to facilitate personal learning that takes students' emotions into account. The system uses a virtual avatar to enhance students' motivation. In this example, one of the interactions with the e-learning system has caused the student to become sad. The figure also shows the context surrounding this interaction; that is, the factors and properties that may have an influence on student's affective states. There are also other factors that may have no influence on the interaction; for instance, the location and other people in the vicinity.

This section is divided into three parts. First, the Affinto ontology that is used as the knowledge base of the proposed platform is detailed. Then, the way that the ontology can be used for analyzing the data extracted from a similar situation (as seen in Figure 1) is described. Finally, the ontology-based platform is described.

A. DESCRIPTION OF THE AFFINTO ONTOLOGY

The Affinto ontology defines the interaction between the user and the system. Figure 2 shows this ontology, developed using the Protégé tool. Various models found in the literature have been taken into account for the design of the ontology. Some user-related models are commonly used in the area of Cognitive Psychology. The system context model is also

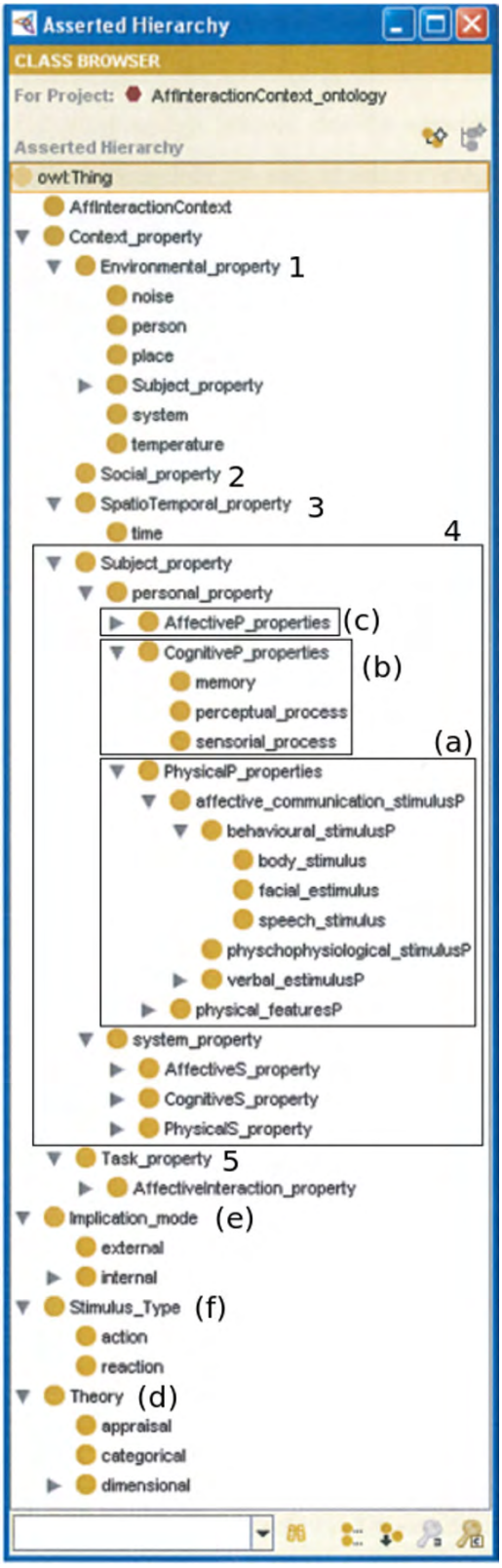

FIGURE 2. Affinto ontology.

defined based on these user models, in a very similar way; i.e. both the user and the system are considered as interlocutors in a given interaction.

Furthermore, the context is often considered as external to the user. However, in this study the user is considered as part of the context (including his/her personal characteristics). Thus, the context covers everything surrounding the affective interaction between the person and the system.

This ontology can be applied to systems that automatically generate interfaces. Thus, the ontology can provide information about both the user and the device's characteristics in order to choose the most suitable multimedia resources. Moreover, it can be applied to multimodal interaction systems. In this case, it can suggest which modality of communication the system should use to interact with a particular user [7].

The Affinto ontology also provides knowledge about affective interactions, because it was found essential to include affective interactions for improving naturalness. That is to say, although the system knows the most appropriate mode of interaction for a given user, if it does not interact in a natural and expressive way with the user, the interaction will remain inadequate for that person.

This context model is the basis of the affective interaction definition because it describes the different factors to both generate and recognize the users' affective states. In addition, this model allows the coherent integration of the involved concepts, since when using the subject context it is possible to describe cognitive processes.

According to the model proposed by Göker and Myrhaug [24], the factors that may have an influence on the interaction are classified into five context elements or properties (see the properties numbered from 1 to 5 in Figure 2). However, in this study, the authors use the *subject context* concept instead of using the *personal context* concept, in order to also include the system contexts and not just the human contexts (see both *personal_property* and *system_property* in box 4 in Figure 2).

These context elements have been modeled in order to represent the knowledge base related to affective interactions. Despite this, the greatest significance has been given to the subject context model (again, box 4 in Figure 2), which includes the subject's physical, cognitive and affective states:

- It is considered that the transmission modes related to human emotions are those proposed by Lang [9] and that they are part of the user's *physical state* in the personal context (see box (a) in Figure 2);
- *Cognitive states* are also included (see box (b) in Figure 2), because cognitive processes take part in the understanding and expression of emotions. From the human point of view, auditory, kinesthetic and visual processes, in addition to language and speech perception and oral processes are considered. From the system point of view, audio extraction, keyboard-mouse input, speech synthesis, video extraction processes, audio parser, video processing and dialog system are included here. In order to do this, the authors use the general model proposed by Wickens [18];

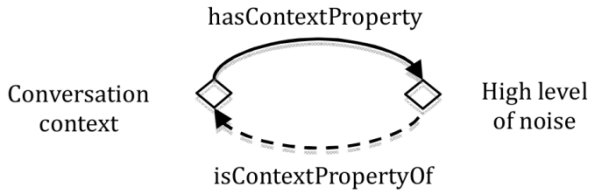

FIGURE 3. The object property called ***hasContextProperty*** that joins the Conversation context instance with the High level of noise instance; and the object property of the inverse function called ***isContextPropertyOf***.

- Finally, *affective states* are also represented (see box (c) in Figure 2), as they have a strong connection with the interaction and communication between people and systems, and even with the physical and cognitive states. The way or tendency users have to experience emotions should be registered and classified by using an appropriate vocabulary. Different emotional theories ([10]–[12]) can be used to represent the same emotion in a distinct way (see element (d) in Figure 2).

The OWL language [31] has been used to develop this ontology. This language allows ontologies to be easily shared, reused, modified and even extended by importing other existing ontologies.

Let us describe the design and the structure of the Affinto ontology. The ontology has five main concepts, as shown in Figure 2: *AffInteractionContext*, *Context_property*, *Implication mode*, *Stimulus_type* and *Theory*. These five concepts are defined below.

AffInteractionContext represents the global context that surrounds the affective interaction between the person and the system. It is composed of several elements or properties, each of which belongs to the *Context_property* class. An object property has been created to define this relationship (see Formula 1).

Formula 1: AffInteractionContext $\rightarrow \exists$ *hasContextProperty some Context_property*

Any property that is within the interaction context is considered as a *Context_property*; each element or detail involved in an interaction, whether a noise, a gesture, a movement, a memory or any stimulus, could affect the subjects and, therefore, their affective states.

The *hasContextProperty* relationship is used to define each property found within the context; e.g. Figure 3 shows a *Conversation context* instance of the *AffInteractionContext* class, which has as a property the *high level of noise* instance, and these are associated with each other using the property called the *hasContextProperty* and its inverse property *isContextPropertyOf*.

Let us now see how the *Context_property* is defined. As mentioned above, each stimulus that exists in an interaction is defined as a *Context_property* and it will belong to at least one context property type: environmental, social, spatio-temporal, task or subject property. One can thus gather information, for instance, about the stimuli that the users have experienced in an interaction or about an environmental factor

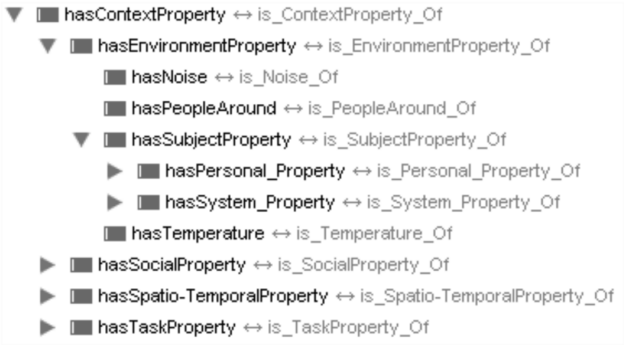

FIGURE 4. Hierarchy of object properties for defining the Context properties.

(such as the environmental temperature) and, also, information about the speech characteristics that a synthesizer should have in a certain situation.

The subject property is in turn also a sub-property of the environmental context, since according to [24]:

"*... This part* [the environmental part] *of the user context captures the entities that surround the user. These entities can for instance be things, services, temperature, light, humidity, noise, or persons. Information (e.g. text, images, movies, sounds) which is accessed by the user in the current user context is all part of the environment context. ...*"

Taking this idea, each subject that is involved in an affective interaction belongs to the environmental context of the other subjects that are also involved in the same interaction.

Based on this classification of the different types of context properties, Affinto also classifies the object properties (see Figure 4). Thus, instead of the *hasContextProperty* of the example in Figure 3, the researchers should use a more specific object, such as *hasEnvironmentProperty* or *hasNoise*.

It is also possible to describe the implication of each subject (either a person or a system). There are a lot of interaction possibilities, for instance: several users sharing experiences with a single system; a given user that has within his/her reach more than one intelligent device; or there could be some people that are not directly involved in an interaction context, but the noise that they are making is affecting an affective interaction. To this end, Affinto uses the Implication mode concept (see the (e) square in Figure 2), which is composed of *external* and *internal* modes of implication. In the case of an internal mode, one can use concepts such as *transmitter* or *receptor* to identify which subject is transmitting a stimulus or who is receiving it (or undergoing some changes in his/her/its affective state). In the case of an external mode, the elements that are not directly involved in the affective interaction, but may influence it can be indicted in the ontology. See Figure 5 for identifying the role of each of the implication modes (where there is a transmitter subject, a receptor subject and some external subjects that are affecting the interaction).

The *hasTheory* object property is used to represent the emotions that compose the affective state; that is, an emotion

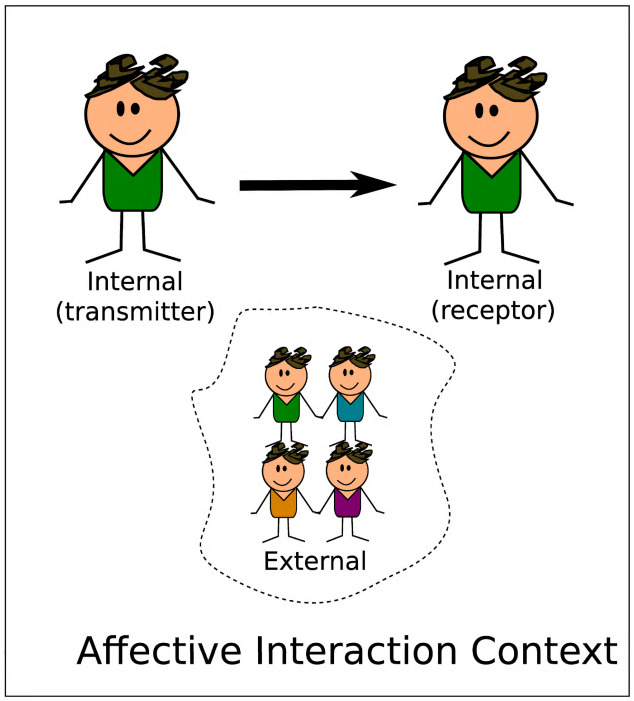

FIGURE 5. An affective interaction between two internal subjects (the transmitter and the receptor) and the influence of some external subjects.

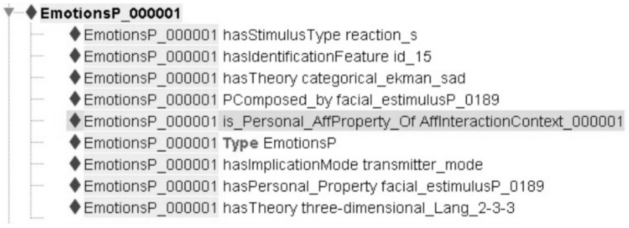

FIGURE 6. The properties related to the EmotionsP_000001 instance (an instance of the Emotion class).

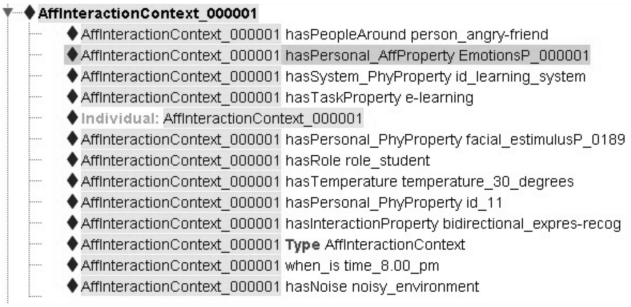

FIGURE 7. The properties of the AffInteractionContext_000001 instance that belongs to the AffInteractionContext class (including EmotionsP_000001).

can be connected with different theories of classifications using this property. So far, three classifications (categorical, dimensional and appraisal) are defined in Affinto. Within each classification, more than one theory can be registered using a *datatype* property called a *Reference*. For example, one can register a stimulus with a *Happy* emotional value and indicate that they have used a categorical theory proposed by Ekman [10] to represent it.

The last main concept of Affinto is *Stimulus_Type* (see element (f) in Figure 2). Emotions are not only influenced by environmental or social factors. Evidently, the stimuli transmitted by his/her interlocutors can also have a great influence on a particular person. Using this concept, one can describe a situation that has occurred and distinguish whether a stimulus emerges as a reaction of another stimulus or not. An interaction is generally a bidirectional process, so it is not enough to analyze the stimuli that a user has transmitted for determining his/her affective state. For instance (returning to the example in Figure 1), the *Other Person*, *Temperature* and *Time* stimuli can be considered as *Action* stimuli, whereas the change of facial features or physiological signals (i.e. the sadness resulting in the user) can be considered as a reaction stimulus. It is thus important to analyze various context properties that may have influenced the interaction in order to understand why the user has reacted in a particular way. Therefore, these concepts help the authors to describe the situations that have given rise to certain affective states in the users. Affinto includes the *hasStimulusType* object property to represent the *Stimulus_Type* of the properties.

B. ANALYSIS OF THE FACTORS INVOLVED GATHERED IN AFFINTO

In order to identify the factors or properties that have affected a given affective state, it is possible to carry out a search for all the uses of that emotion (see Figure 6).

One can also identify the *Affective Interaction Context* instance corresponding to this emotion (i.e. the context where the emotion has arisen) to analyze all the properties and stimuli involved in the interaction and in which way they are involved. One can use the *is_PersonalAffProperty_Of* inverse function to identify this instance (see *AffInteractionContext_000001*, highlighted in Figure 6). Figure 7 shows the properties involved in the *Affective Interaction Context* of the scenario of Figure 1.

Regarding the environmental context, one can see that certain factors (such as the environmental temperature or another person who was an annoyed friend) have affected the interaction. One can also see that time may have an influence on the person and the noise in the system, both of which are part of the environmental context of this interaction. One can include additional information such as the role (a student) of the person in this interaction or the task (e-learning) that both subjects of the interaction are performing.

C. DESCRIPTION OF THE ONTOLOGY-BASED PLATFORM

As previously mentioned, the use of the ontologies allows the authors to gather information as a knowledge base and to analyze this information in order to recognize, interpret, and generate affective states by means of the use of different resources or computational applications. Therefore, based

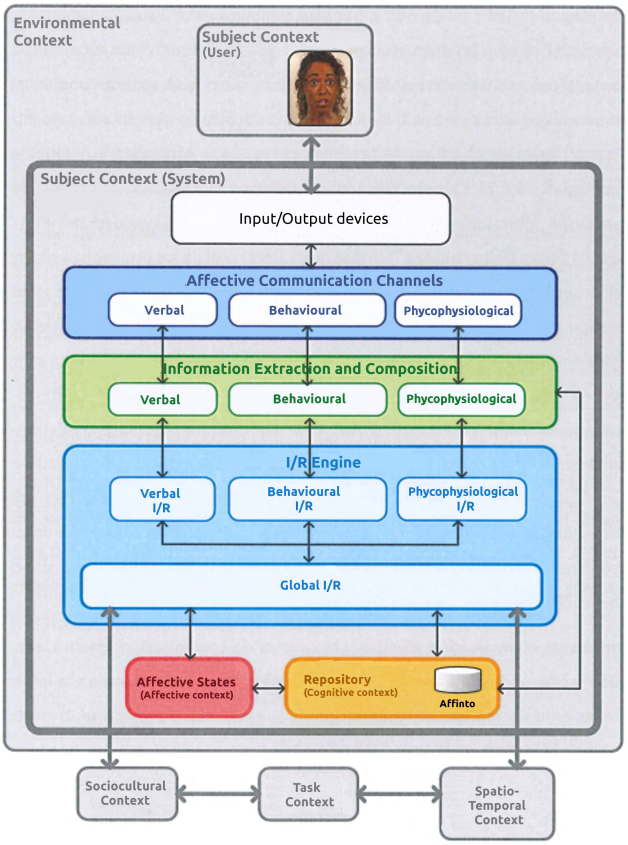

FIGURE 8. **The ontology-based platform for developing affective interaction systems.**

on the Affinto ontology, a platform has been created for making the development of these kinds of applications easier (see Figure 8).

The main goal of the ontology-based platform proposed in this paper is to support the development of resources that need to take emotion transmission into account in the interactions between users and systems.

This platform is composed of several modules. Within the environmental context there are two subjects (the person and the system) and the other context types are also included (the sociocultural, task and spatio-temporal contexts).

Depending on the functionality of the interaction system to be developed (i.e. whether an emotional recognition process and/or an emotional synthesis process is needed), the procedure to be performed with these modules will be different (depending on the direction of the communication between the user and the system):

1) PHASE ONE

In the *Affective recognition process* (when the user transmits information to the system) the procedure with the modules is the following (more details about the procedure are presented in next section):

Step 1.1 (Input/Output Devices Module): Depending on the communication channels that the system uses, the information will be transmitted to the corresponding input devices.

Step 1.2: Then, the *Information Extraction and Composition module* extracts the necessary data from the message (e.g. facial or voice features).

Step 1.3 (Interpretation/Response (I/R) Engine Module): the process corresponding to the communication channel used is performed in order to analyze the extracted data. For instance, some data mining techniques are applied to the extracted features for estimating the *Affective State* of the message.

Step 1.4: For analyzing these data and estimating the Affective State, the *I/R Engine* uses the information gathered in the *Repository* that is composed mainly of the Affinto ontology.

Step 1.5: In the case that the interaction is bidirectional (i.e. the system has to generate a response to the user), suitable mechanisms (e.g. a dialogue system) have to be used. The *I/R Engine* also manages this process.

If the system is multimodal more than one communication channel (each with its own percentage) must be taken into account in this recognition process.

2) PHASE TWO

In the *affective synthesis process* (when the system generates information to be sent to the user):

Step 2.1 (I/R Engine Module): As previously mentioned in step 1.5 of the recognition process, a message has to be generated in order to interact with the user. Therefore, the necessary data have to be collected in order to compose a suitable message. This module can use the Affinto ontology to identify suitable information or features that a synthesizer should use in a particular message.

Step 2.2: Having identified the suitable data, the *Information Extraction and Composition module* composes the message. That is, the system has to include the previously obtained information as parameters of the synthesizer.

Step 2.3: The message is redirected to the corresponding *communication channel* and finally is transmitted to the user through the corresponding *output device*.

These two procedures are explained in greater detail in the next section, as a multimodal interaction system, based on the proposed platform, has been developed. The aim of the creation of this system is to validate the development of emotional resources guided by the proposed platform.

IV. VALIDATION OF THE ONTOLOGY-BASED PLATFORM BY MEANS OF AN EMPIRICAL STUDY

The validation process of the platform was performed in two main steps. In the first step, an interaction system was developed for only one communication modality; specifically the verbal modality (written text). The results of this validation were published in [6]. In the second step, a multimodal interaction system was developed. In this section, it will first be shown how the ontology-based platform guided the development of this multimodal interaction system; specifically it is an affective conversational system called AFFIN. Then, the empirical study with the created conversational system is explained.

Bearing in mind that it is a conversational system, it includes recognition, interpretation and synthesis processes. To make all these processes possible, the system uses the ontology, and the information stored in it, by means of software not developed by the authors of the ontology. Using external software reinforces the usefulness of the ontology and the platform for creating affective resources and/or systems.

A. AFFIN: A MULTIMODAL AFFECTIVE CONVERSATIONAL SYSTEM

Conversational systems, also known as dialogue systems, are intelligent interfaces that allow users to interact with them. They generally use one of the most common communication modalities of human beings (speech) and represent a major advance in human-computer interaction technology.

These systems also integrate technologies such as automatic voice recognition, natural language processing, dialogue management, speech synthesis, etc. [32]. In order to validate the platform as a supporting tool for the development of Affective Computing applications and in turn, to validate the Affinto ontology as a knowledge base for these applications, the authors have developed AFFIN: a multimodal conversational system for text and speech that is able to recognize, interpret and generate emotions. This system integrates the technologies mentioned above. Regarding the emotional recognition system of AFFIN, the process performed is a personalized process. That is, in order to identify the emotion transmitted by a given user, the classifier that performs the recognition uses data previously stored by this same user. In this way, if enough information for each user is collected, the results obtained are more accurate than the results that are achieved using a general corpus composed of numerous users' characteristics.

Figure 9 shows the architecture created for the development of the AFFIN system. As can be seen, the design of this architecture is based on the proposed ontology-based platform and some of the modules have been developed for achieving the objectives of the conversational system. It can be seen that the channels that the system uses for interacting with the person are verbal (for the transmission of verbal information by means of the written language) and speech (for the transmission of paralinguistic features through speech).

The media resources that the system offers through its interface and its communication channels should not avoid information about emotions. To this end, by using the Affinto ontology, the system can extract information about the required features for emotionally enriching the interface, as well as the stimulus to be emitted to the user. Conversely, the system can also extract information about certain user's communication modality characteristics in order to recognize the user's affective state by using the information previously gathered in the ontology.

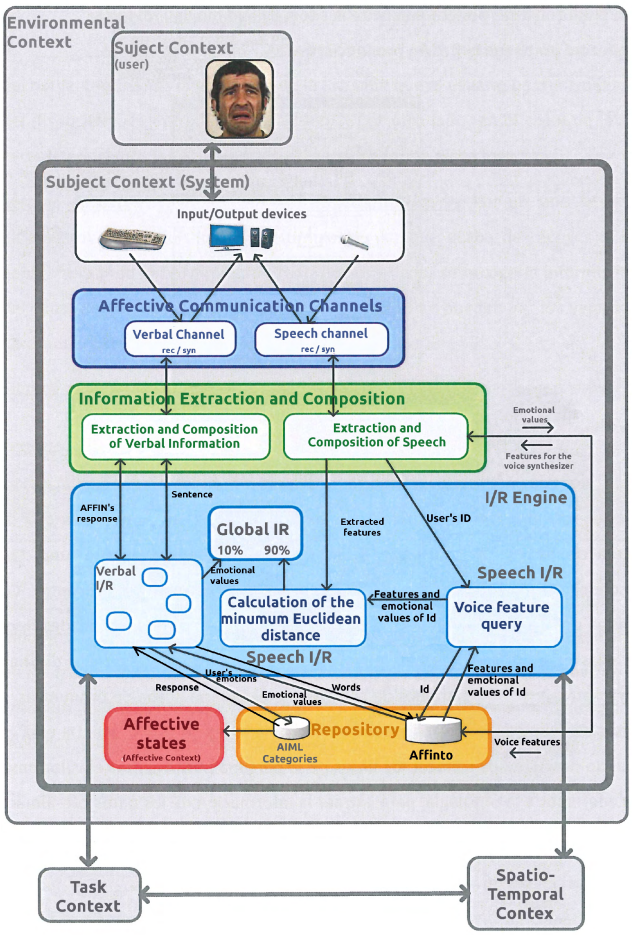

FIGURE 9. Architecture of the AFFIN system.

Let us again divide the interaction between the user and the system into two parts: the affective recognition process and the affective synthesis process.

1) PHASE ONE

Affective recognition process with AFFIN. Two general steps are distinguished, one for each communication channel: (a) text and (b) speech.

Step 1.1 (Text-Based Affective Recognizer Process): Regarding the text-based affective recognizer (see the verbal channel in Figure 9), a method based on affective dictionaries has been selected. In this case, the ANEW affective dictionary [33] is used. Each word in ANEW has an emotional value represented by means of three dimensions: valence, arousal and dominance. Each of these words has been registered in the Affinto ontology as an interaction, together with its emotional value [6]. Once the authors have all this information in the ontology, the process performed by the text-based affective recognizer is as follows.

First, the *Verbal Channel module* receives the text through the *input device*, the microphone. AFFIN uses a voice recognizer called Sphinx-4 [34] for extracting the words transmitted by the user. The *Verbal Channel module* then sends

the message to the *Information Extraction and Composition module*. This module performs a syntactic analysis of the message to enable the *Verbal I/R module* to label nouns, adverbs, adjectives and verbs with emotional values. A parser created by The Stanford Natural Language Processing group [35] is used to do this. The parser also detects the words that are negation dependent in order to invert their emotional values. In the *I/R Engine*, the emotional values of these words are thus determined by using the *Repository* (i.e. the Affinto ontology) to obtain the average emotional value of a given text. In this way, the authors obtain the emotional value of the text transmitted by the user.

It is known that the non-verbal information transmitted has more importance than the verbal information in human conversations [1]. For that reason, when interpreting the emotion transmitted by the user, the system (see the *Global I/R module* in Figure 9) will take 10% of the value obtained by the text-based affective recognizer, according to Mehrabian's estimation.

Step 1.2 (Speech-Based Affective Recognizer Process): The remaining 90% is obtained from the speech-based affective recognizer. Regarding this speech recognizer, the process is the following: as in the text recognizer, the speech recognizer also uses the Sphinx-4 tool. In this case, Sphinx-4 makes a recording of the transmitted voice in the *Speech Channel*. The *Information Extraction and composition module* (in this process, the module corresponding to speech) then extracts the paralinguistic features of the voice using a tool called Praat [36]. Using this tool, the AFFIN system extracts eleven features from the user's voice. These eleven features are grouped into three categories: (1) The tone of voice or Pitch (also known as Fundamental Frequency) - average value, maximum, minimum and the standard deviation; (2) Voice intensity or volume - average value, maximum, minimum and the standard deviation; (3) Formants (intensity peaks in the spectrum of a sound) - with F1 as the lowest frequency formant, followed by F2 and F3.

Once the system has performed the extraction of these features, the *I/R Engine module* interprets this information. The information stored in the Affinto ontology is used for this purpose. This information is collected when the users carry out a training process; i.e. the first time they use the conversational system. It is also possible to gather the interactions performed after the training process. Each of these interactions is stored and labeled with the emotional values transmitted by the user. There are several techniques to recognize emotions in speech [37], [38]. In this study, the K-Nearest Neighbors (K-NN) algorithm [39] has been applied. Therefore, the authors have the recently extracted set of features from a given user's interaction and also the set of features corresponding to several interactions performed by this user, including the emotional values. The *Speech I/R module* can therefore obtain the affective state of this user in the interaction, based only on the speech features, by applying this algorithm.

2) PHASE TWO (AFFECTIVE SYNTHESIS PROCESS WITH AFFIN)

After recognizing the affective state of the user, the system generates a response. To this end, it has to interpret the user's message together with its affective value. The *I/R Engine module* uses a dialogue system developed by the ALICE project [40] to do this. This system uses the AIML Markup Language with an interpreter for this language. The interpreter has some AIML categories in the *repository* for creating the message depending on the text input, but more categories have been created for this validation. The main objective for creating new AIML categories is to include emotions as input information and to select a suitable response based on these emotions. Moreover, emotional information is also included in the responses of the interpreter. Therefore, two communication channels are distinguished again:

Step 2.1 (Text-Based Affective Synthesis Process): The above-mentioned response message is sent to the *Verbal Communication Channel*.

Step 2.2 (Speech-Based Affective Synthesis Process): A suitable set of features is sent to the speech channel, so that a voice synthesizer (called FreeTTS [41]) can generate an emotional utterance through the *output device*; i.e. the speakers. In order to identify the features corresponding to the emotion that the synthesizer has to transmit, AFFIN once again uses the ontology.

Having generated the synthesized voice message, AFFIN is now ready to receive the next message from the user.

B. THE EMPIRICAL STUDY WITH THE AFFIN SYSTEM

In the following sub-sections, the empirical study for validating the AFFIN system (performed with some experimental subjects) is presented. The main objective of this study is to prove that the use of the ontology-based platform facilitates the development of Affective Computing applications, even using externally developed software. The study is also useful for proving that the Affinto ontology serves as a knowledge base for these types of applications.

1) PARTICIPANTS IN THE EXPERIMENT

14 volunteers participated in the experiment: 9 men (average age of 32.22; sd = 9.00; age range = 24-53) and 5 women (average age of 32; sd = 8.07; age range = 26-47). They were asked to indicate their level in English. Five of them responded *good* and the remaining nine responded *acceptable*.

Considering that it can be quite difficult to represent the emotions transmitted by participants via the three dimensions, an inter-judge agreement test was performed. Three evaluators had to listen to all the recordings from the training process. In this test, a correlation coefficient called Kendall's Tau-b was measured in order to compare what each participant says that he/she expressed with the evaluators' opinions on the same recordings. The result of the test for the Dominance dimension was that the correlation obtained between

four judges (a participant and three evaluators) was not significant for any of the 14 participants. It can be deduced that the expression of the emotions through this dimension is very difficult for both participants and judges. For the other two dimensions, the judges had more agreement in the case of Valence than in the case of Arousal.

Since the Kendall's Tau-b coefficients were quite low, the data with less agreement were discarded and only those that showed high and significant correlation ($p < 0.05$ level for bilateral prediction) were considered valid. 6 of the 14 participants were therefore discarded and the validation was performed with the results of the remaining 8 participants. In this way, the authors ensured that the participants performed the training exercise correctly.

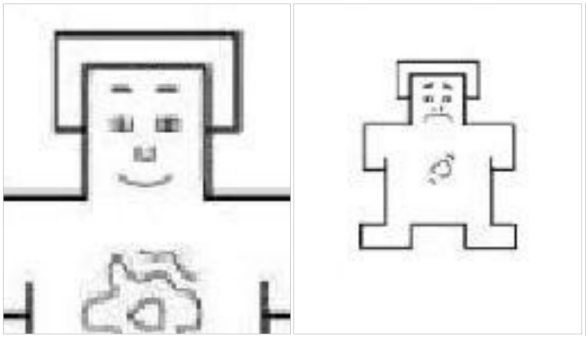

FIGURE 10. Examples of two images that integrate the three SAM scales into a single image. The values that they represent are: (a) (7, 5, 9) and (b) (1, 3, 3) respectively.

2) MATERIAL AND TOOLS

The AFFIN system was set up in order to start training and adapt the interaction to each user and his/her characteristics. For the study, the results of the AFFIN recognizer were analyzed.

With relation to external resources, IAPS (International Affective Picture System) images [42] were used to induce emotions in the participants.

Some JSGF [43] grammars for the Sphinx-4 recognizer and some AIML categories for the dialogue system were also created for controlling the experiment.

In addition, the dimensional theory was chosen for representing the emotions of the users. To this end, the SAM measurement tool (see section about Related Work and Models) was used, but applying some modifications to it.

The most relevant is that in the user interface of the system, where the results of the recognizer are shown: instead of using a different image to display emotions for each of the three scales, the values of the three dimensions are shown integrated into a single image, together with the exact values of the system's result. For example, if the system wants to represent an emotion with a '7' in the Valence scale, a '5' in the Arousal scale and a '9' in the Dominance scale, instead of using the three scales and a score for each one of the scales, the system shows the image presented in Figure 10 (a). The objective of this change is to be able to easily and directly see the representation of an emotion in a single image.

3) DESIGN OF THE EXPERIMENT

The experiment was divided into four phases. The first three are for correctly performing the training of AFFIN and the last is for performing the affective recognition in real time.

Since the main objective of the experiment is to check that the system is able to conduct conversations with the participants and is able to interpret the emotions that they transmit, in the training phases the participants have to indicate the emotion that they really wanted to transmit, in order to check the accuracy (although this is not the main objective of the experiment). Thus, the recognizer has a base for recognizing the emotions in interactions and the authors can also ensure that the data stored in the ontology for subsequent recognitions are correct.

The design of the experiment is an intra-subject design; i.e. all the subjects or participants have to perform the four phases. The subjects transmitted 38 valid utterances in total. The language used in the experiment was English.

4) PROCEDURE FOR THE EXPERIMENT

The experimenter met each participant individually in one room. First, the participant received general and specific instructions for performing the experiment and he/she had to complete a demographic questionnaire. Afterwards, he/she began the session itself. Each session lasted about an hour.

The process that the participants followed is described below, phase-by-phase:

a: PHASE ONE (BASIC TRAINING)

The participant has to pronounce 18 sentences with the emotions indicated in the interface. The authors thus obtain voice features corresponding to different emotions (using dimensional representation, such as $\langle \text{valence} = 1; \text{arousal} = 1; \text{dominance} = 5 \rangle$).

b: PHASE TWO (TOUCHDOWN)

The system starts a conversation with a simple question or a greeting and the participant responds. Depending on the response (message and emotion transmitted), AFFIN continues with different questions. In this phase, the subjects can express their own emotions, but if AFFIN does not correctly recognize the emotions they have to correct these values. In this phase, the participant has to pronounce 5 sentences in total.

c: PHASE THREE (RESPONSES EXPRESSED BY DIFFERENT EMOTIONS)

In this phase, the questions have been created for being answered by the participant expressing specific emotions (based on the dimensional values). In this case, the participant has to choose one of the three proposed answers (the one with

which he/she most identifies). To help participants to feel and express these emotions two IAPS system images are shown for each sentence, which are intended to induce emotions in the participants. Each participant utters 9 sentences in this third phase.

d: PHASE FOUR (AFFECTIVE RECOGNITION BY MEANS OF REAL TIME CLASSIFICATION)

As in the third phase, the participant has to choose one of the three proposed answers. However, in this case the emotion that he/she has to express is freely chosen (which is more natural). He/she does not have to correct the emotional values recognized by the system, because this is not a training phase. However, he/she has to indicate the real emotion expressed for later comparison with the affective recognizer's result. The participant has to transmit 6 utterances in this fourth phase. With this, the experiment is concluded.

5) RESULTS OF THE EXPERIMENT

In order to carry out the analysis of the results obtained in the experiment, an assessment of inter-judge agreement was performed. Kendall's Tau-b correlation coefficients were also calculated. To this end, the correlation between the results obtained from the recognition system of AFFIN in the fourth phase and the emotional values indicated by the participants was analyzed.

The results prove that for most of the participants the correlation is positive, but not significant. This may be due to the low number of samples ($N = 6$). To enlarge the sample size, the Kendall's Tau-be coefficient has been calculated over the entire data set for all participants in this fourth phase (although the whole set of data is evaluated at the same time, this method evaluates the agreement of various judges over the same data). Thus, the sample size becomes $N = 48$ (6 interactions for each of the 8 participants).

The highest coefficient was obtained in the Dominance dimension, Kendall's Tau-b $= 0.368$, $N = 48$, $p = 0.02$; then, in the Valence dimension, Kendall's Tau-b $= 0.329$, $N = 48$, $p = 0.04$; and finally, in the Arousal dimension, Kendall's Tau-b $= 0.208$, $N = 48$, $p = 0.06$.

Table 1 shows the average error for each participant (ranging from 0 to 8, since the SAM scale uses values from 1 to 9) and the percentages of accuracy of the emotion recognition. All these data are based on the emotional dimensions for the third phase of the training and for the real- time classification of the fourth phase.

Some of these percentages of accuracy are not very high, but even in humans it is nearly impossible to reach 100%. One of the reasons could be that the language used in the experiment is not the mother tongue of the participants. Thus, the conversation held with AFFIN was not totally natural and the participants were unable to express the emotions that they wanted to express. In addition, in some cases they probably unwittingly indicated emotions that were not the true emotions, thus incorrectly training the system. Evidence of this is the assessment of inter-judge agreement performed by three evaluators. Analyzing the Kendall's Tau-b correlation coefficients, it is considered that it was difficult for them to express the emotions that they wanted to for the three dimensions.

TABLE 1. Error differences and accuracy percentages for emotion recognition, during and after training. Val. = Valence; Aro. = Arousal; Dom. = Dominance; P1-P8 = Participant identifications; % = Accuracy percentages for emotion recognition.

Error differences of each participant	Training			Real-time classification		
	Val.	Aro.	Dom.	Val.	Aro.	Dom.
P1	3.45	3.63	1.35	1.36	0.86	1.90
P2	1.69	2.61	0.70	2.11	2.04	0.43
P3	1.31	2.14	0.49	2.24	2.01	0.59
P4	2.55	4.25	0.69	2.87	3.41	1.32
P5	1.59	1.81	0.86	1.29	2.96	1.32
P6	1.43	1.64	0.47	0.95	1.56	0.71
P7	3.05	4.26	0.69	2.50	2.57	1.79
P8	1.53	1.72	0.28	1.04	1.98	0.68
Average error	2.08	2.76	0.69	1.79	2.17	1.09
%	74.00	65.50	91.38	77.83	72.88	86.38

Another reason could be that the training does not incorporate a large volume of data. Most recognition techniques use large databases to classify the features obtained from the users and thus interpret the emotion expressed by them. In this case, the recognition process performed is a personalized process where the participant trains his/her own behavior and the way of expressing emotion to the system. Thereby, it is possible to adapt the system to the person. However, in a single session the system cannot store a sufficient amount of data to accurately interpret the conversations held with the participant.

6) DISCUSSION ABOUT THE EXPERIMENT

The first assessment of inter-judge agreement was useful for determining which participants had correctly performed the training process and, thus, discarding those who did not present high and significant correlation (by means of the measurement of Kendall's Tau-b coefficients at $p < 0.05$ level for a bilateral prediction).

A K-NN inductive learning algorithm was used for each participant performing the training process, in order to associate the data with different emotions. As previously presented in Table 1, there was positive correlation between the values obtained by the recognition system and the values given by participants in the real-time classification of the fourth phase. Moreover, this correlation was significant in the case of Valence and Dominance (at level $p < 0.05$, bilateral prediction). The correlation for Arousal was not significant, but with a p value that was not very high ($p = 0.06$).

Regarding the values obtained in the training phase, the results improved in the fourth phase (see Table 1). The authors can deduce that the results are improved by adding more information to the recognizer, with the exception of the

dimension Dominance. Although the results in this dimension were the most optimal, in the fourth phase they worsened instead of improving. Moreover, the correlation was not significant for the Dominance dimension in the first assessment of inter-judge agreement with the three evaluators. This may be because the users are usually accustomed to using categories such as joy, sadness, fear etc. to represent emotions. Since the theory used in this validation was the dimensional theory, the participants found it more difficult to indicate the emotion that they transmitted in each utterance.

For future tests, authors plan including professional actors as participants in the experiments, thinking that they should be able to simulate desired emotions more accurately.

In spite of all this, and considering that the size of the data set of the training base is not very large, the authors can say that the results were quite good and satisfactory.

V. CONCLUSION

In this paper, the authors have presented an ontology (Affinto) that defines affective states and, also, the interaction between humans and systems. Using this ontology, it is possible to evaluate a situation that has given rise to certain emotions and the stimuli or properties derived from them. In this way, it is possible to create patterns for the automatic recognition of emotions in human beings and to motivate users with appropriate responses.

Furthermore, Affinto's description of affective interactions has made possible the creation of a platform for developing Affective Computing applications for several areas, such as e-education, telemedicine, and so forth [6] and [8]. In turn, the ontology has served as a knowledge base for an affective multimodal conversational system (AFFIN).

Several users participated in the study to validate this system. Its main objective was to analyze the affective reaction caused in participants by the conversation directed by AFFIN and by the images displayed in order to induce emotions. To analyze these reactions the system extracts the paralinguistic parameters from participants' voices and the verbal information from their messages. In this way, the system performs a recognition process and obtains an estimate of the subject's affective state. The experimental results show that there is positive correlation between the values obtained by the recognition system and the values indicated by participants. It must be highlighted that the main goal of this paper is not to present a precise system that recognizes human emotions during a conversation with the users, but rather to present an ontology that serves as knowledge base and, based on this ontology, a platform that serves as guide for developing Affective Computing systems.

Due to the use of an ontology-based approach, other intelligent agents could also access the information stored in the Affinto ontology, by using it as an information repository or retrieving information in a semantic manner. In addition, the platform based on Affinto could also serve as a guide for the development of other affective resources and/or applications.

Moreover, both the ontology and the platform are modular. In this case, the text and speech modalities have been developed, but other modalities can also be included. The authors are currently working on the analysis of physiological signals (such as GSR, ECG or EMG) [44], in order to detect behavioral patterns and recognize emotions based on these signals. After creating these patterns, the authors will be able to include the features of these physiological signals in Affinto and combine them with the other communication modalities.

ACKNOWLEDGMENT

The authors would like to thank the participants of the experiments that were performed to validate the system.

REFERENCES

[1] A. Mehrabian, *Silent Messages*. Belmont, CA, USA: Wadsworth, 1971.
[2] R. W. Picard, *Affective Computing*. Cambridge, MA, USA: MIT Press, 1997.
[3] *Affective Computing Research Group of the Massachusetts Institute of Technology (MIT)*. Accessed: Feb. 14, 2019. [Online]. Available: http://affect.media.mit.edu
[4] K. Höök, "Affective computing," in *The Encyclopedia of Human-Computer Interaction*, 2nd ed, M. Soegaard and R. F. Dam, Eds. Aarhus, Denmark: Interaction Design Foundation, 2012.
[5] A. Gómez-Pérez, M. Fernandez-Lopez, and O. Corcho, "Ontological engineering: With examples from the areas of knowledge management," in *e-Commerce and The Semantic Web*. New York, NY, USA: Springer-Verlag, 2003.
[6] I. Cearreta and N. Garay-Vitoria, "Applying the *Affinto* ontology to develop a text-based emotional conversation system," in *Proc. INTERACT*, Lisbon, Portugal, vol. 4, 2011, pp. 479–482.
[7] I. Cearreta and N. Garay-Vitoria, "Toward adapting interactions by considering user emotions and capabilities," in *Proc. HCI*, Orlando, FL, USA, vol. 14, Jul. 2011, pp. 525–534.
[8] J. M. López, R. Gil, R. García, I. Cearreta, and N. Garay, "Towards an ontology for describing emotions," in *Proc. WSKS*, Athens, Greece, Sep. 2008, pp. 96–104.
[9] P. J. Lang, "The emotion probe: Studies of motivation and attention," *Amer. Psycholog.*, vol. 50, no. 5, pp. 372–385, 1995.
[10] P. Ekman, "Expression and nature of emotion," in *Approaches to Emotion*, K. Scherer and P. Ekman Eds. Mahwah, NJ, USA: Erlbaum, 1984, pp. 319–343.
[11] H. Schlosberg, "Three dimensions of emotion," *Psychol. Rev.*, vol. 61, no. 2, pp. 81–88, 1954.
[12] K. R. Scherer, "Appraisal theory," in *Handbook Cognition Emotion*, T. Dalgleish and M. J. Power, Eds. New York, NY, USA: Wiley, 1999, pp. 637–663.
[13] O. Pierre-Yves, "The production and recognition of emotions in speech: Features and algorithms," *Int. J. Hum.-Comput. Stud.*, vol. 59, nos. 1–2, pp. 157–183, 2003.
[14] A. Tellegen, "Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report," in *Anxiety and the Anxiety Disorders*, A. H. Tuma and J. D. Maser, Eds. Hillsdale, NJ, USA: Lawrence Erlbaum, 1985, pp. 681–706.
[15] M. Schröder, R. Cowie, E. Douglas-Cowie, M. Westerdijk, and S. Gielen, "Acoustic correlates of emotion dimensions in view of speech synthesis," in *Proc. 7th Eur. Conf. Speech Commun. Technol.*, Aalborg, Denmark, vol. 1, Sep. 2001, pp. 87–90.
[16] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *J. Behav. Therapy Exp. Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.
[17] N. Malandrakis, A. Potamianos, E. Iosif, and S. Narayanan, "Distributional semantic models for affective text analysis," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 11, pp. 2379–2392, Nov. 2013.
[18] C. D. Wickens, *Engineering Psychology & Human Performance*. New York, NY, USA: Harper Collins, 1992.

[19] J. J. Cañas, L. Salmerón, and P. Gámez, "Human factor in human-computer interaction," (in Spanish), in *Introductory Course on Human-Computer Interaction*, J. Lorés, Ed. Tallahassee, FL, USA: AIPO, 2001.
[20] (2015). *W3C Multimodal Interaction Working Group Charter*. Accessed: Feb. 28, 2019. [Online]. Available: https://www.w3.org/2013/10/mmi-charter
[21] W3C Recommendation. (2014). *EmotionML (Emotion Markup Language) 1.0*. Accessed: Feb. 28, 2019. [Online]. Available: https://www.w3.org/TR/emotionml/
[22] W3C Recommendation. (2009). *EMMA (Extensible MultiModal Annotation Markup Language) 1.0*. Accessed: Feb. 28, 2019. [Online]. Available: https://www.w3.org/TR/emma/
[23] D. A. Dahl, *Multimodal Interaction With W3C Standards: Toward Natural User Interfaces to Everything*. Cham, Switzerland: Springer, 2017.
[24] A. Göker and H. I. Myrhaug, "User context and personalization," in *Proc. 6th Eur. Conf. Case Based Reasoning (ECCBR)*, Aberdeen, U.K., Sep. 2002, pp. 1-7.
[25] K.-I. Benta, A. Rarău, and M. Cremene, "Ontology based affective context representation," in *Proc. Euro Amer. Conf. Telematics Inf. Syst.*, Faro, Portugal, May 2007, p. 46.
[26] S. Bhatia, M. Hayat, and R. Goecke, "A multimodal system to characterise melancholia: Cascaded bag of words approach," in *Proc. 19th ACM Int. Conf. Multimodal Interact. (ICMI)*, Glasgow, Scotland, Nov. 2017, pp. 274–280.
[27] E. Cambria, I. Hupont, A. Hussain, E. Cerezo, and S. Baldassarri, "Sentic avatar: Multimodal affective conversational agent with common sense," in *Toward Autonomous, Adaptive, and Context-Aware Multimodal Interfaces. Theoretical and Practical Issues* (Lecture Notes in Computer Science), vol. 6456, A. Esposito, A. M. Esposito, R. Martone, V. C. Müller, and G. Scarpetta, Eds. Berlin, Germany: Springer-Verlag, 2011, pp. 81–95.
[28] N. Reitano. (2018). Digital Emotions: The Potential and Issues of Affective Computing Systems. MsC Digital Interactive Media, University of Dublin, Dublin, Ireland, Accessed: Feb. 27, 2019. [Online]. Available: https://scss.tcd.ie/publications/theses/diss/2018/TCD-SCSS-DISSERTATION-2018-071.pdf
[29] A. Sengupta, A. Dasgupta, A. Chaudhuri, A. George, A. Routray, and R. Guha, "A multimodal system for assessing alertness levels due to cognitive loading," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 7, pp. 1037–1046, Jul. 2017.
[30] N. Thompson and T. J. McGill, "Affective stack—A model for affective computing application development," *J. Softw.*, vol. 10, no. 8, pp. 919–930, 2015.
[31] *Web Ontology Language (OWL)—W3C Recommendation*. Accessed: Feb. 10, 2004. [Online]. Available: https://www.w3.org/TR/2004/REC-owl-features-20040210/
[32] R. López-Cózar, and M. Gea, "Ubiquitous dialog system for educational environments," (in Spanish), in *Proc. of Interacción*, Lleida, Spain, May 2004, pp. 101–108.
[33] M. Bradley and P. Lang, "Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings," Univ. Florida, Center Res. Psychophysiol., Gainesville, FL, USA, Tech. Rep. C-1, 1999.
[34] *Sphinx-4. (n.d.)*. Accessed: Feb. 14, 2019. [Online]. Available: http://cmusphinx.sourceforge.net/wiki/tutorialsphinx4/
[35] *Stanford Natural Language Processing Group. (n.d.)*. Accessed: Feb. 14, 2019. [Online]. Available: http://nlp.stanford.edu/software/lex-parser.shtml
[36] *Praat: Doing Phonetics by Computer. (n.d.)*. Accessed: Feb. 14, 2019. [Online]. Available: http://www.fon.hum.uva.nl/praat/
[37] K. Dai, H. J. Fell, and J. MacAuslan, "Recognizing emotion in speech using neural networks," in *Proc. IASTED Int. Conf. Telehealth/Assistive Technol.*, Baltimore, MD, USA, vol. 18, 2008, pp. 31–36.
[38] D. Gharavian, M. Sheikhan, A. Nazerieh, and S. Garoucy, "Speech emotion recognition using FCBF feature selection method and GA-optimized fuzzy ARTMAP neural network," *Neural Comput. Appl.*, vol. 21, no. 8, pp. 2115–2126, 2012.
[39] K. M. Ting, "Common issues in instance-based and naive-Bayesian classifiers," Ph.D. dissertation, Dept. Comput. Sci., Univ. Sidney, Sydney NSW, Australia, 1995.
[40] *ALICE AI Foundation. (n.d.)*. Accessed: Jan. 8, 2018. [Online]. Available: http://www.alicebot.org/about.html
[41] *FreeTTS: A Speech Synthesizer Written Entirely in the JavaTM Programming Language. (n.d.)*. Accessed: Feb. 14, 2019. [Online]. Available: http://freetts.sourceforge.net/docs/index.php
[42] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "International affective picture system (IAPS): Affective ratings of pictures and instruction manual," Dept. Comput. Sci., Univ. Florida, Gainesville, FL, USA, Tech. Rep. A-8, 2008.
[43] *JSGF Grammar Format. (n.d.)*. Accessed: Feb. 14, 2019. [Online]. Available: https://www.w3.org/TR/jsgf/
[44] F. Canento, A. Fred, H. Silva, H. Gamboa, and A. Lourenço, "Multimodal biosignal sensor data handling for emotion recognition," in *Proc. SENSORS*, Limerick, Ireland, vol. 31, Oct. 2011, pp. 647–650.

NESTOR GARAY-VITORIA received the M.S. degree in informatics and the Ph.D. degree in computer science from the University of the Basque Country (UPV/EHU), Donostia, in 1990 and 2000, respectively, where he is currently an Associate Professor with Computer Architecture and Technology Department. He was the first Academic Director of the Master's Program on Assistive Technology for Personal Autonomy given in UPV/EHU. His research interest includes human-computer interaction for special needs in Egokituz Research Group. He is also interested on computer science education at all educational levels. He received the Extraordinary Award for the Ph.D. Dissertation, in 2002.

IDOIA CEARRETA received the B.Sc. and Ph.D. degrees in computing from the University of the Basque Country (UPV/EHU), Donostia, in 2004 and 2010, respectively. She was with the Egokituz Research Group. She currently teaches at pre-university levels.

EDURNE LARRAZA-MENDILUZE received the M.S. and Ph.D. degrees in computer science from the University of the Basque Country (UPV/EHU), Donostia, Spain, in 1999 and 2015, respectively, where she is currently an Assistant Professor with the Computer Architecture and Technology Department. Her research interest includes computer science education at all educational levels. She is a member of the ADIAN Research Group.