Sentic Avatar: Multimodal Affective Conversational Agent with Common Sense

Erik Cambria[1], Isabelle Hupont[2],
Amir Hussain[1], Eva Cerezo[3], and Sandra Baldassarri[3]

[1] University of Stirling, Stirling, UK
[2] Aragon Institute of Technology, Zaragoza, Spain
[3] University of Zaragoza, Zaragoza, Spain
{eca,ahu}@cs.stir.ac.uk, ihupont@ita.es, {sandra,ecerezo}@unizar.es
http://cs.stir.ac.uk/~eca/sentics

Abstract. The capability of perceiving and expressing emotions through different modalities is a key issue for the enhancement of human-computer interaction. In this paper we present a novel architecture for the development of intelligent multimodal affective interfaces. It is based on the integration of Sentic Computing, a new opinion mining and sentiment analysis paradigm based on AI and Semantic Web techniques, with a facial emotional classifier and Maxine, a powerful multimodal animation engine for managing virtual agents and 3D scenarios. One of the main distinguishing features of the system is that it does not simply perform emotional classification in terms of a set of discrete emotional labels but it operates in a continuous 2D emotional space, enabling the integration of the different affective extraction modules in a simple and scalable way.

Keywords: AI, Sentic Computing, NLP, Facial Expression Analysis, Sentiment Analysis, Multimodal Affective HCI, Conversational Agents.

1 Introduction

Emotions are a fundamental component in human experience, cognition, perception, learning and communication. A user interface cannot be considered really intelligent unless it is also capable of perceiving and expressing emotions. For this reason, affect sensing and recognition from multiple modalities is getting a more and more popular research field for the enhancement of human-computer interaction (HCI).

In this paper we present a novel architecture for integrating a process for reasoning by analogy over affective knowledge, a facial emotional classifier and a multimodal engine for managing 3D virtual scenarios and characters.

The structure of the paper is the following: Section 2 presents the state of the art of multimodal affective HCI, Section 3 briefly describes the proposed architecture, Section 4 explains in detail how the affect recognition and integration are performed and, eventually, Section 5 comprises concluding remarks and a description of future work.

A. Esposito et al. (Eds.): COST 2102 Int. Training School 2010, LNCS 6456, pp. 81–95, 2011.

2 Multimodal Affective HCI

Human computer intelligent interaction is an emerging field aimed at providing natural ways for humans to use computers as aids. It is argued that for a computer to be able to interact with humans it needs to have the communication skills of humans. One of these skills is the affective aspect of communication, which is recognized to be a crucial part of human intelligence and has been argued to be more fundamental in human behaviour and success in social life than intellect [1][2]. Emotions influence cognition, and therefore intelligence, especially when it involves social decision-making and interaction.

The latest scientific findings indicate that emotions play an essential role in decision-making, perception, learning and more. Most of the past research on affect sensing has considered each sense such as vision, hearing and touch in isolation. However, natural human-human interaction is multimodal: we communicate through speech and use body language (posture, facial expressions, gaze) to express emotion, mood, attitude, and attention.

Affect recognition from multiple modalities has a short historical background and is still in its first stage [3]. It was not till 1998 that computer scientists attempted to use multiple modalities for recognition of emotions/affective states [4]. The results were promising: using multiple modalities improved the overall recognition accuracy helping the systems function in a more efficient and reliable way. Following the findings in psychology, which suggested that the most significant channel for judging emotional cues of humans is the visual channel of face and body [5], a number of works combine facial expressions and body gestures for affect sensing [6][7][8]. Other approaches combine different biological information such as brain signals or skin conductivity for affect sensing [9][10].

However this research makes use of a single information channel, i.e. a single type of computer input device, and, therefore, must assume the reliability on this channel. For that reason, the trend in recent works is to consider and combine affective information coming from different channels. That way, eventual changes on the reliability of the different information channels are considered.

Recent literature on multimodal affect sensing is focused on the fusion of data coming from the visual and audio channels. Most of those works make use of the visual channel for body gesture recognition [11] or facial expression classification [12] and the audio channel to analyze non-linguistic audio cues such as laughters [13], coughs [14] or cries [15]. However, very few works fuse information coming from the visual channel with linguistic-based (speech contents) audio affect sensing. With all these new areas of research in affect sensing, a number of challenges have arisen. In particular, the synchronization and fusion of the information coming from different channels is a big problem to solve.

Previous studies fused emotional information either at a decision-level, in which the outputs of the unimodal systems are integrated by the use of suitable expert criteria [16], or at a feature-level, in which the data from both modalities are combined before classification [17]. In any case, the choice of fusion strategy depends on the targeted application.

Accordingly, all available multimodal recognizers have designed and/or used ad-hoc solutions for fusing information coming from multiple modalities but cannot accept new modalities without re-defining the whole system. In summary, there is not a general consensus when fusing multiple modalities and systems' scalability is not possible.

3 Overview of the System

The architecture proposed (illustrated in Fig. 1) is based on the multimodal animation engine Maxine [18], and it consists of four main modules: Perception, Affective Analysis, Deliberative/Generative and Motor module.

The Perception module simply consists of the hardware necessary to gather the multimodal information from the user i.e. keyboard, microphone and webcam. The Affective Analysis module aims to infer the user's affective state from the different inputs and integrate it. The Deliberative/Generative module is in charge of processing the extracted emotional information to manage the virtual agent's decisions and reactions, which are finally generated by the Motor module. In this paper we focus on the presentation of the affective sensing part of the system, explained in detail in Section 4.

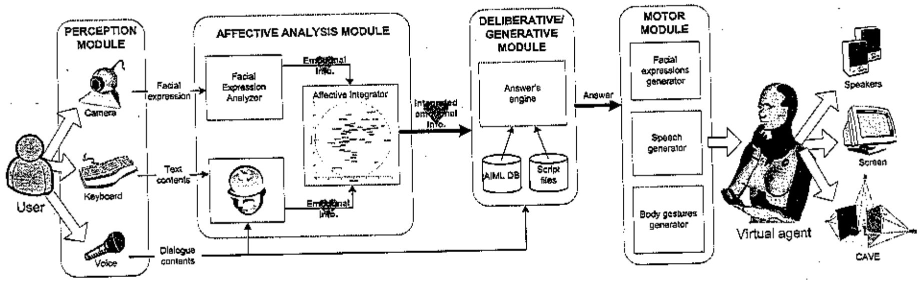


Fig. 1. Sentic Avatar's architecture

4 Affective Analysis Module

The Affective Analysis Module is in charge of extracting emotions from the textual, vocal and video inputs and integrate them.

It consists of three main parts: Sentic Computing, for inferring emotions from typed-in text and speech-to-text converted contents (Section 4.1), Facial Expression Analyzer, for extracting affective information from video (Section 4.2), and Affective Integrator, for integrating the outputs coming from the two previous modules (Section 4.3).

4.1 Sentic Computing

Sentic Computing [19] is a new opinion mining and sentiment analysis paradigm which exploits AI and Semantic Web techniques to better recognize, interpret and process opinions and sentiments in natural language text.

In Sentic Computing, whose term derives from the Latin 'sentire' (the root of words such as sentiment and sensation) and 'sense' (intended as common sense), the analysis of text is not based on statistical learning models but rather on common sense reasoning tools [20] and domain-specific ontologies [21]. Differently from keyword spotting [22][23][24], lexical affinity [25][26] and statistical [27][28][29] approaches, which generally requires large inputs and thus cannot appraise texts with satisfactory granularity, Sentic Computing enables the analysis of documents not only on the page or paragraph-level but also on the sentence-level.

AffectiveSpace. AffectiveSpace [30] is a multi-dimensional vector space built from ConceptNet [31], a semantic network of common sense knowledge, and WordNet-Affect, a linguistic resource for the lexical representation of affective knowledge [32]. The blend [33] of these two resources yields a new dataset in which common sense and affective knowledge coexist i.e. a 14,301 $\times$ 117,365 matrix whose rows are concepts (e.g. 'dog' or 'bake cake'), whose columns are either common sense and affective features (e.g. 'isA-pet' or 'hasEmotion-joy'), and whose values indicate truth values of assertions.

In this knowledge base each concept is represented by a vector in the space of possible features whose values are positive for features that produce an assertion of positive valence (e.g. 'a penguin is a bird'), negative for features that produce an assertion of negative valence (e.g. 'a penguin cannot fly') and zero when nothing is known about the assertion. The degree of similarity between two concepts, then, is the dot product between their rows in the blended matrix.

The value of such a dot product increases whenever two concepts are described with the same feature and decreases when features that are negations of each other describe them. When performed on the blended matrix, however, these dot products have very high dimensionality (as many dimensions as there are features) and are difficult to work with. In order to approximate these dot products in a useful way, we project all of the concepts from the space of features into a space with many fewer dimensions i.e. we reduce the dimensionality of the matrix by means of principal component analysis (PCA).

In particular, we perform truncated singular value decomposition (TSVD) [34] in order to obtain a new matrix that forms a low-rank approximation of the original data and represents, for the Eckart–Young theorem [35], the best estimation of the original data in the Frobenius norm sense. By exploiting the information sharing property of truncated SVD, concepts with the same affective valence are likely to have similar features – that is, concepts conveying the same emotion tend to fall near each other in AffectiveSpace. Concept similarity does not depend on their absolute positions in the vector space, but rather on the angle they make with the origin.

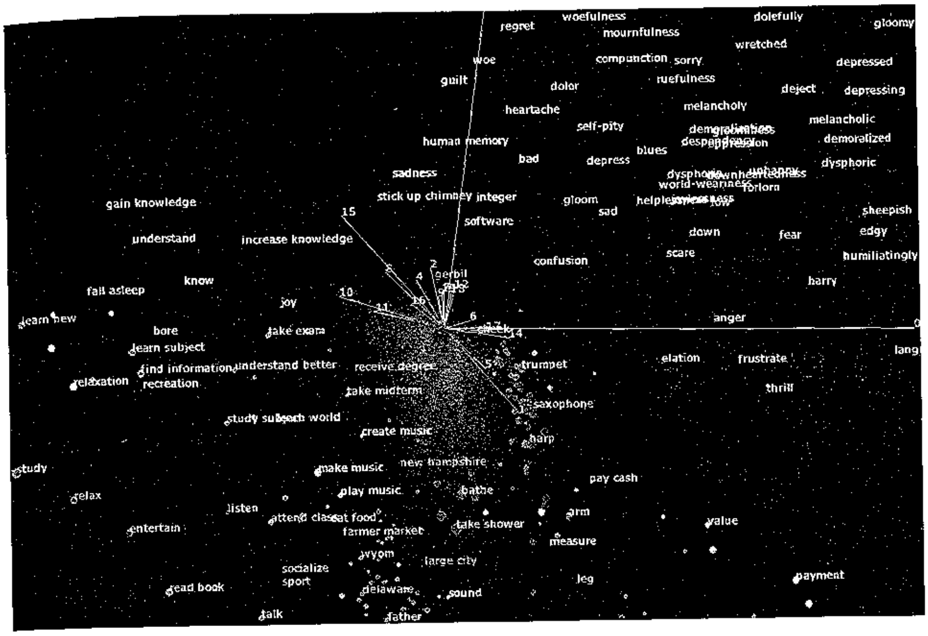
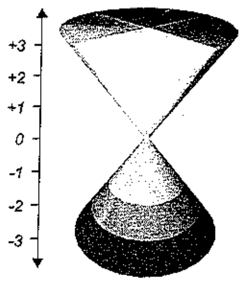


Fig. 2. A sketch of AffectiveSpace

For example we can find concepts such as 'beautiful day', 'birthday party', 'laugh' and 'make person happy' very close in direction in the vector space, while concepts like 'sick', 'feel guilty', 'be laid off' and 'shed tear' are found in a completely different direction (nearly opposite with respect to the center of the space). If we choose to discard all but the first 100 principal components, common sense concepts and emotions are represented by vectors of 100 coordinates: these coordinates can be seen as describing concepts in terms of 'eigenmoods' that form the axes of AffectiveSpace i.e. the basis e_0,...,e_{99} of the vector space. For example, the most significant eigenmood, e_0, represents concepts with positive affective valence. That is, the larger a concept's component in the e_0 direction is, the more affectively positive it is likely to be. Concepts with negative e_0 components, then, are likely to have negative affective valence.

The Hourglass of Emotions. The Hourglass of Emotions is an affective categorization model developed starting from Plutchik's studies on human emotions [36]. In the model (illustrated in Fig. 3) sentiments are reorganized around four independent dimensions whose different levels of activation make up the total emotional state of the mind. The Hourglass of Emotions, in fact, is based on the idea that the mind is made of different independent resources and that emotional states result from turning some set of these resources on and turning another set of them off [37].

The model is particularly useful to recognize, understand and express emotions in the context of human-computer interaction (HCI). In the Hourglass of



	Pleasantness	Attention	Sensitivity	Aptitude
+3	ecstasy	vigilance	rage	admiration
+2	joy	anticipation	anger	trust
+1	serenity	interest	annoyance	acceptance
0	—	—	—	—
-1	pensiveness	distraction	apprehension	boredom
-2	sadness	surprise	fear	disgust
-3	grief	amazement	terror	loathing

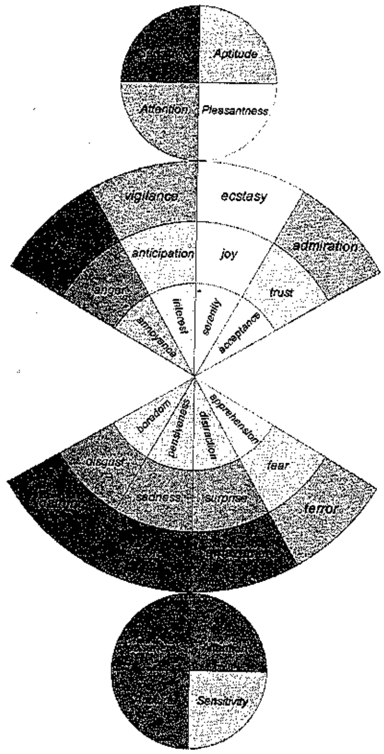


Fig. 3. The Hourglass of Emotions

Emotions, in fact, affective states are not classified, as often happens in the field of emotion analysis, into a few basic categories, but rather into four concomitant but independent dimensions – Pleasantness, Attention, Sensitivity and Aptitude – in order to understand how much respectively:

1. the user is happy with the service provided
2. the user is interested in the information supplied
3. the user is comfortable with the interface
4. the user is disposed to use the application

Each affective dimension is characterized by six levels of activation that determine the intensity of the expressed/perceived emotion, for a total of 24 labels specifying 'elementary emotions'. The concomitance of the different affective dimensions makes possible the generation of 'compound emotions' such as *love*, given by the combination of *joy* and *trust*, or *disappointment*, given by the concomitance of *surprise* and *sadness*.

Sentics Extraction Process. The text contents typed-in by the user and the dialogue contents, traslated into text using Loquendo automatic speech recognition (ASR) [38], go through a Natural Language Processing (NLP) module,

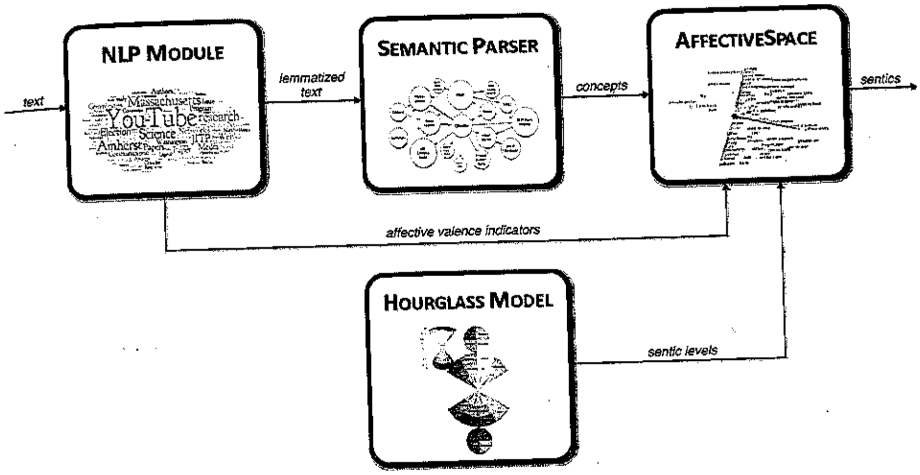


Fig. 4. Sentics Extraction Process

which performs a first skim of the document, a Semantic Parser, whose aim is to extract concepts from the processed text, and eventually AffectiveSpace, for the inference of concepts' affective valence (Fig. 4).

In particular, the NLP module interprets all the affective valence indicators usually contained in text such as special punctuation, complete upper-case words, onomatopoeic repetitions, exclamation words, degree adverbs or emoticons, and it ultimately lemmatizes text.

The Semantic Parser exploits a concept n-gram model extracted from ConceptNet graph structure in order to deconstruct text into common sense concepts. Once concepts are retrieved from the lemmatized text, the parser also provides, for each of these, the relative frequency, valence and status i.e. the concept's occurrence in the text, its positive or negative connotation, and the degree of intensity with which the concept is expressed.

The AffectiveSpace module, eventually, projects the retrieved concepts into the vector space clustered (using a k-means approach) wrt the Hourglass model, and infers the affective valence of these according to the positions they occupy in the multi-dimensional space.

Therefore, the outputs of the Sentics Extraction Process are four dimensional vectors, called 'sentic vectors', which specify the emotional charge carried by each concept in terms of Pleasantness, Attention, Sensitivity and Aptitude. This information is ultimately exploited to produce emotional labels which are given as inputs to the Affective Integrator, as shown in the example below:

Text: Yesterday was a BEAUTIFUL day!
I couldn't find the food I was after and there was pretty bad weather
but I bought a new dress and a lot of Christmas presents.

<Concept: 'yesterday'>
<Concept: 'beautiful day'++>

<*Concept: !'find food'*>
<*Concept: 'bad weather'*-->
<*Concept: 'buy new dress'*>
<*Concept: 'buy christmas present'++*>

Sentics: [1.608, 0.571, 0.0, 2.489]
Moods: joy and interest
Polarity: 0.51

4.2 Facial Expression Analyzer

The Facial Expression Analyzer achieves an automatic classification of the shown facial expressions into discrete emotional categories. It is able to classify the user's emotion in terms of Ekman's six universal emotions (*fear*, *sadness*, *joy*, *disgust*, *surprise* and *anger*) [39] plus *neutral*, giving a membership confidence value to each emotional category.

The face modeling selected as input for the Facial Expression Analyzer follows a feature-based approach: the inputs are a set of facial distances and angles calculated from feature points of the mouth, eyebrows and eyes. In fact, the inputs are the variations of these angles and distances with respect to the neutral face. The points are obtained thanks to a real-time facial feature tracking program [40].

Fig. 5(a) shows the correspondence of these points with those defined by the MPEG4 standard. The set of parameters obtained from these points is shown in Fig. 5(b). In order to make the distance values consistent (independently of the scale of the image, the distance to the camera, etc.) and independent of the expression, all the distances are normalized with respect to the distance between the eyes i.e. the MPEG4 Facial Animation Parameter Unit (FAPU), also called ESo. The choice of angles provides a size invariant classification and saves the effort of normalization. As regards the classification process itself, the system intelligently combines the outputs of 5 different classifiers simultaneously. In this way, the overall risk of making a poor selection with a given classifier for a given input is reduced.

The classifier combination chosen follows a weighted majority voting strategy, where the voted weights are assigned depending on the performance of each classifier for each emotion. In order to select the best classifiers to combine, the Waikado Environment for Knowledge Analysis (Weka) tool was used [41]. This provides a collection of machine learning algorithms for data mining tasks. From this collection, five classifiers were selected after tuning: RIPPER, Multilayer Perceptron, SVM, Naive Bayes and C4.5.

The selection was based on their widespread use as well as on the individual performance of their Weka implementation. To train the classifiers and evaluate the performance of the system, two different facial emotion databases were used: the FGNET database [42] that provides video sequences of 19 different Caucasian people, and the MMI Facial Expression Database [43] that holds 1280 videos of 43 different subjects from different races (Caucasian, Asian and Arabic). Both databases show Ekman's six universal emotions plus *neutral*.

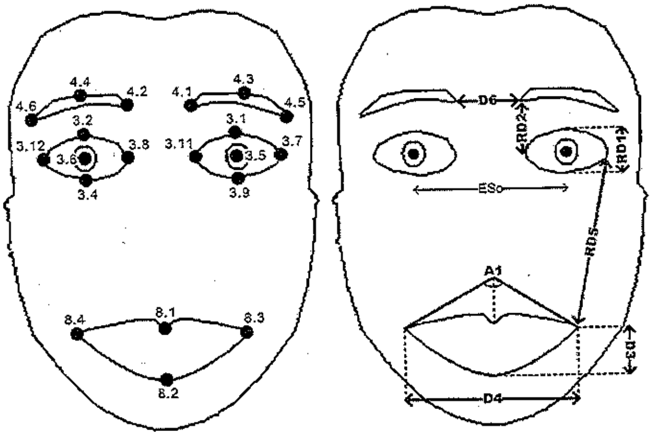


Fig. 5. (a) Tracked facial feature points according to MPEG4 standard and (b) corresponding facial parameters

Table 1. Confusion matrix obtained combining the five classifiers

classified as	disgust	joy	anger	fear	sadness	neutral	surprise
disgust	**79.41%**	0%	2.39%	18.20%	0%	0%	0%
joy	4.77%	**95.23%**	0%	0%	0%	0%	0%
anger	19.20%	0%	**74.07%**	0%	3.75%	2.98%	0%
fear	9.05%	0%	0%	**62.96%**	8.53%	0%	19.46%
sadness	0.32%	0.20%	1.68%	0%	**30.00%**	67.80%	0%
neutral	0%	0%	1.00%	2.90%	4.10%	**92.00%**	0%
surprise	0%	0%	0%	11.23%	0%	4.33%	**84.44%**

A new database has been built for testing this work with a total of 1500 static frames carefully selected from the apex of the video sequences from the FG-NET and MMI databases. The results obtained when applying the strategy explained previously to combine the scores of the five classifiers are shown in the form of confusion matrix in Table 1 (results have been obtained with a 10-fold cross-validation test over the 1500 database images).

As can be observed, the success rates for *neutral*, *joy* and *surprise* are very high (84.44%–95.23%). However, the system tends to confuse *disgust* with *fear*, *anger* with *disgust* and *fear* with *surprise*; therefore, the performances for those emotions are slightly worse. The lowest result of our classification is for *sadness*: it is confused with *neutral* on 67.80% of occasions, due to the similarity of the facial expressions. Confusion between these pairs of emotions occurs frequently in the literature and for this reason many classification works do not consider some of them. Nevertheless, the results can be considered positive as two incompatible emotions (such as *sadness* and *joy* or *fear* and *anger*) are confused on less

Table 2. Confusion matrix obtained after considering human assessment

classified as	disgust	joy	anger	fear	sadness	neutral	surprise
disgust	**84.24%**	0%	2.34%	13.42%	0%	0%	0%
joy	4.77%	**95.23%**	0%	0%	0%	0%	0%
anger	15.49%	0%	**77.78%**	0%	3.75%	2.98%	0%
fear	1.12%	0%	0%	**92.59%**	2.06%	0%	4.23%
sadness	0.32%	0.20%	1.68%	0%	**66.67%**	31.13%	0%
neutral	0%	0%	0%	0.88%	1.12%	**98.00%**	0%
surprise	0%	0%	0%	6.86%	0%	2.03%	**91.11%**

than 0.2% of occasions. Another relevant aspect to be taken into account when evaluating the results is human opinion.

The labels provided in the database for training classifiers correspond to the real emotions felt by users although they do not necessarily have to coincide with the perceptions other human beings may have about the facial expressions shown. Undertaking this kind of study is very important when dealing with human-computer interaction, since the system is proved to work in a similar way to the human brain. In order to take into account the human factor in the evaluation of the results, 60 persons were told to classify the 1500 images of the database in terms of emotions. As a result, each one of the frames was classified by 10 different persons in 5 sessions of 50 images.

The Kappa statistic obtained from raters annotations is equal to 0.74 (calculated following the formula proposed in [44]), which indicates an adequate inter-rater agreement in the emotional images annotation. With this information, the evaluation of the results was repeated: the recognition was marked as good if the decision was consistent with that reached by the majority of the human assessors. The results (confusion matrix) of considering users' assessment are shown in Table 2. As can be seen, the success ratios have considerably increased. Therefore, it can be concluded that the confusions of the algorithms go in the same direction as those of the users: our classification strategy is consistent with human classification.

4.3 Affective Integrator

The Sentics Extraction Process outputs a list of sentic vectors which represents an emotional analysis of text and dialogue contents in terms of Pleasantness, Attention, Sensitivity and Aptitude while the Facial Expression Analyzer provides an affective evaluation of video contents in terms of Ekman's six universal emotions. Some researchers, such as Whissell [45] and Plutchik [46], consider emotions as a continuous 2D space whose dimensions are evaluation and activation. The evaluation dimension measures how a human feels, from positive to negative. The activation dimension measures whether humans are more or less likely to take some action under the emotional state, from active to passive. To overcome the problem of the integration of the affective information coming from the Sentic Extraction

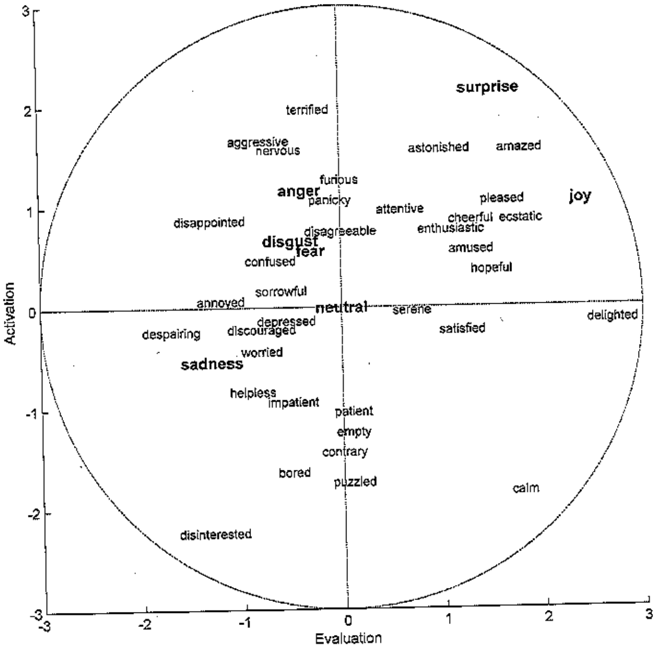


Fig. 6. Whissell's model

Process and the Facial Expression Analyzer, the continuous 2D description of affect is considered.

Bi-dimensional representations of affect are attractive mainly because they provide a way of describing emotional states that is more tractable than using words. This is of particular importance when dealing with naturalistic data, where a wide range of emotional states occurs. Similarly, they are much more able to deal with non discrete emotions and variations in emotional states over time, since in such cases changing from one universal emotion label to another would not make much sense in real life scenarios.

In her study, Cynthia Whissell assigns a pair of values activation-evaluation to each of the approximately 9000 words with affective connotations that make up her Dictionary of Affect in Language. Fig. 6 shows the position of some of these words in the activation-evaluation space.

The emotion-related words corresponding to each one of Ekman's six emotions plus *neutral* and to the levels of Pleasantness, Attention, Sensitivity and Aptitude have a specific location in the Whissell space. Thanks to this, in our work the output information of the Sentic Extraction Process and the labels provided by the Facial Expression Analyzer can be mapped in the Whissell space: a pair of values activation-evaluation can be calculated from the obtained labels, and hence concurrently visualized and compared in the 2D space (Fig. 7).

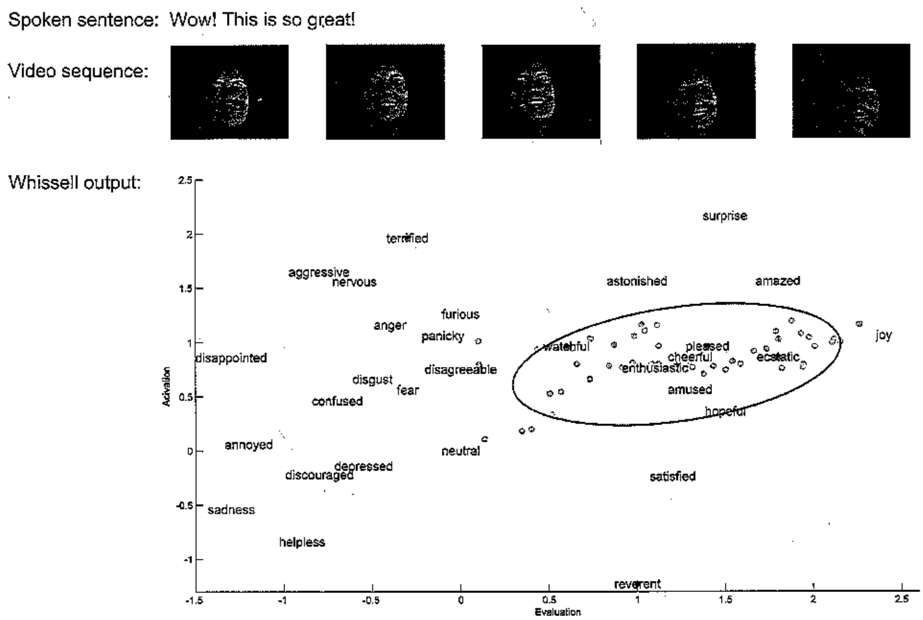


Fig. 7. Example of Affective Analysis Module output: integration of the extracted emotional information (from audio and video) into the Whissell space

In particular, the process of affective integration is achieved through the following three steps:

1. Each one of the emotional labels inferred by the Sentics Extraction Process from the video spoken sentence is mapped as a 2D point on to the Whissell space.
2. In the same way, the Facial Expression Analyzer outputs the user's emotion in terms of Ekman's six universal emotions (plus *neutral*), giving a membership confidence value to each emotional category. The mapping of its output in the Whissell space is carried out considering each of Ekman's six basic emotions plus *neutral* as 2D weighted points in the activation-evaluation space, where the weights are assigned depending on the confidence value obtained for each emotion in the classification process. The final detected emotion is calculated as the centre of mass of the seven weighted points in the Whissell space. That way, the Facial Expression analyzer outputs one emotional location in the Whissell space per frame of the studied video sequence.
3. Finally, the whole set of 2D activation-evaluation points obtained from both the Sentics Extraction Process and the Facial Expression Analyzer is fitted to the Minimum Volume Ellipsoid (MVE) that better covers the shape of the set of extracted points. The MVE is calculated following the algorithm described by Kumar and Yildrim [47]. The final emotional information outputted by affective analysis module for the whole video sequence is given by the x-y coordinates of the centre of that MVE.

5 Conclusion and Future Work

This paper describes a novel architecture for multimodal affective interaction with conversational agents. The proposed system recognizes user's affective state through two different modalities: AffectiveSpace, a process for reasoning by analogy and association over common sense and affective knowledge able to extract affective information in terms of Pleasantness, Attention, Sensitivity and Aptitude from spoken words (dialogue contents), and a Facial Expression Analyzer that classifies the shown user's facial expression in terms of the six Ekman's universal emotions (plus *neutral*).

The emotional information coming from these two different channels is integrated in a simple manner thanks to Whissell's 2D activation-evaluation space, where all the extracted emotional labels are mapped onto x-y locations of that space. The benefits of using Whissell 2D representation are, on the one hand, that the final output of the system does not simply provide a classification in terms of a set of emotionally discrete labels but goes further by extending the emotional information over an infinite range of intermediate emotions and, on the other hand, its capability for improving the overall recognition accuracy helping the system function in a more reliable way.

Furthermore, it opens the door to the integration of new emotional extraction modules in the future (e.g. modules that study user's gestures, gaze, mouse-clicks or keyboard use for affective recognition) in a simple and scalable fashion.

References

1. Vesterinen, E.: Affective Computing. In: Digital Media Research Seminar, Helsinki (2001)
2. Pantic, M.: Affective Computing. Encyclopedia of Multimedia Technology and Networking. Hershy 1, 8–14, Idea Group Reference (2005)
3. Gunes, M., Gunes, H., Piccardi, M., Pantic, M.: From the Lab to the Real World: Affect Recognition Using Multiple Cues and Modalities. In: Affective Computing: Focus on Emotion Expression, Synthesis and Recognition, pp. 185–218 (2008)
4. Riseberg, J., Klein, J., Fernandez, R., Picard, R.: Frustrating the User on Purpose: Using Biosignals in a Pilot Study to Detect the User's Emotional State. In: CHI, Los Angeles (1998)
5. Ambady, N., Rosenthal, R.: Thin Slices of Expressive Behavior as Predictors of Interpersonal Consequences: a Meta-Analysis. Psychological Bulletin 11(2), 256–274 (1992)
6. Camurri, A., Mazzarino, B., Volpe, G.: Analysis of Expressive Gesture: The EyesWeb Expressive Gesture Processing Library. In: Gesture Workshop, Genova (2003)
7. Gunes, H., Piccardi. M.: Bi-Modal Emotion Recognition from Expressive Face and Body Gestures. Network and Computer Applications 30(4), 1334–1345 (2007)
8. Karpouzis, K., Caridakis, G., Kessous, L., Amir, N., Raouzaiou, A., Malatesta, L., Kollias, S.: Modeling Naturalistic Affective States Via Facial, Vocal and Bodily Expressions Recognition. In: Huang, T.S., Nijholt, A., Pantic, M., Pentland, A. (eds.) ICMI/IJCAI Workshops 2007. LNCS (LNAI), vol. 4451, pp. 92–116. Springer, Heidelberg (2007)

9. Pun, T., Alecu, T., Chanel, G., Kronegg, J., Voloshynovskiy, S.: Brain-Computer Interaction Research at the Computer Vision and Multimedia Laboratory. IEEE Trans. on Neural Systems and Rehabilitation Engineering 14(2), 210–213 (2006)
10. Burleson, W., Picard, R., Perlin, K., Lippincott, J.: A Platform for Affective Agent Research. In: International Conference on Autonomous Agents and Multiagent Systems, New York (2004)
11. Petridis, S., Pantic, M.: Audiovisual Discrimination between Laughter and Speech. In: ICASSP, Las Vegas (2008)
12. Valstar, M., Gunes, H., Pantic, M.: How to Distinguish Posed from Spontaneous Smiles Using Geometric Features. In: ICMI, Nagoya (2007)
13. Truong, K., Van Leeuwen, D.: Automatic Discrimination Between Laughter and Speech. Speech Communication 49, 144–158 (2007)
14. Matos, S., Birring, S., Pavord, I., Evans, D.: Detection of Cough Signals in Continuous Audio Recordings Using HMM. IEEE Trans. on Biomedical Engineering 53(6), 1078–1083 (2006)
15. Pal, P., Iyer, A., Yantorno, R.: Emotion Detection from Infant Facial Expressions and Cries. In: Intl Conf. Acoustics, Speech and Signal Processing (2006)
16. Jong-Tae, J., Sang-Wook, S., Kwang-Eun, K., Kwee-Bo, S.: Emotion Recognition Method Based on Multimodal Sensor Fusion Algorithm. In: ISIS, Sokcho-City (2007)
17. Shan, C., Gong, S., McOwan, P.: Beyond Facial Expressions: Learning Human Emotion from Body Gestures. In: BMVC, Warwick (2007)
18. Baldassarri, S., Cerezo, E., Seron, F.: Maxine: a Platform for Embodied Animated Agents. Computers and Graphics 32(4), 430–437 (2008)
19. Cambria, E., Hussain, A., Havasi, C., Eckl, C.: Sentic Computing: Exploitation of Common Sense for the Development of Emotion-Sensitive Systems. In: Esposito, A., Campbell, N., Vogel, C., Hussain, A., Nijholt, A. (eds.) COST 2102. LNCS, vol. 5967, pp. 153–161. Springer, Heidelberg (2010)
20. Cambria, E., Hussain, A., Havasi, C., Eckl, C.: Common Sense Computing: From the Society of Mind to Digital Intuition and Beyond. In: Fierrez, J., Ortega-Garcia, J., Esposito, A., Drygajlo, A., Faundez-Zanuy, M. (eds.) BioID MultiComm 2009. LNCS, vol. 5707, pp. 252–259. Springer, Heidelberg (2009)
21. Cambria, E., Grassi, M., Hussain, A., Havasi, C.: Sentic Computing for Social Media Marketing. In: Multimedia Tools and Applications. Springer, Heidelberg (to appear, 2010)
22. Elliott, C.: The Affective Reasoner: A Process Model of Emotions in a Multi-Agent System. The Institute for the Learning Sciences, Technical Report No. 32 (1992)
23. Ortony, A., Clore, G., Collins, A.: The Cognitive Structure of Emotions. Cambridge University Press, New York (1988)
24. Wiebe, J., Wilson, T., Claire, C.: Annotating Expressions of Opinions and Emotions in Language. Language Resources and Evaluation 39(2), 165–210 (2005)
25. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In: HLT-EMNLP, Vancouver (2005)
26. Somasundaran, S., Wiebe, J., Ruppenhofer, J.: Discourse Level Opinion Interpretation. In: COLING, Manchester (2008)
27. Hu, M., Liu, B.: Mining Opinion Features in Customer Reviews. In: AAAI, San Jose (2004)
28. Pang, B., Lee, L.: Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with Respect to Rating Scales. In: ACL, Ann Arbor (2005)

29. Abbasi, A., Chen, H., Salem, A.: Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums. ACM Transactions on Information Systems 26(3), 1–34 (2008)
30. Cambria, E., Hussain, A., Havasi, C., Eckl, C.: AffectiveSpace: Blending Common Sense and Affective Knowledge to Perform Emotive Reasoning. In: WOMSA at CAEPIA, Seville (2009)
31. Havasi, C., Speer, R., Alonso, J.: ConceptNet 3: a Flexible, Multilingual Semantic Network for Common Sense Knowledge. In: RANLP, Borovets (2007)
32. Strapparava, C., Valitutti, A.: WordNet-Affect: an Affective Extension of WordNet. In: LREC, Lisbon (2004)
33. Havasi, C., Speer, R., Pustejovsky, J., Lieberman, H.: Digital Intuition: Applying Common Sense Using Dimensionality Reduction. IEEE Intelligent Systems 24(4), 24–35 (2009)
34. Wall, M., Rechtsteiner, A., Rocha, L.: Singular Value Decomposition and Principal Component Analysis. In: Berrar, D., et al. (eds.) A Practical Approach to Microarray Data Analysis, pp. 91–109. Kluwer, Norwell (2003)
35. Eckart, C., Young, G.: The Approximation of One Matrix by Another of Lower Rank. Psychometrika 1(3), 211–218 (1936)
36. Plutchik, R.: The Nature of Emotions. American Scientist 89(4), 344–350 (2001)
37. Minsky, M.: The Emotion Machine. Simon and Schuster, New York (2006)
38. Loquendo Audio Speech Recognition, `http://www.loquendo.com`
39. Ekman, P., Dalgleish, T., Power, M.: Handbook of Cognition and Emotion. Wiley, Chihester (1999)
40. Cerezo, E., Hupont, I., Manresa, C., Varona, J., Baldassarri, S., Perales, F., Seron, F.: Real-Time Facial Expression Recognition for Natural Interaction. In: Martí, J., Benedí, J.M., Mendonça, A.M., Serrat, J. (eds.) IbPRIA 2007. LNCS, vol. 4478, pp. 40–47. Springer, Heidelberg (2007)
41. Witten, I., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, San Francisco (2005)
42. Wallhoff, F.: Facial Expressions and Emotion Database. Technische Universitat Munchen (2006)
43. Pantic, M., Valstar, M., Rademaker, R., Maat, L.: Web-Based Database for Facial Expression Analysis. In: ICME, Singapore (2005)
44. Siegel, S., Castellan, N.: Nonparametric Statistics for the Social Siences. McGraw-Hill, New York (1988)
45. Whissell, C.: The Dictionary of Affect in Language. Emotion: Theory, Research and Experience, The Measurement of Emotions 4, 113–131 (1989)
46. Plutchik, R.: Emotion: a Psychoevolutionary Synthesis. Harper and Row, New York (1980)
47. Kumar, P., Yildirim, E.: Minimum-Volume Enclosing Ellipsoids and Core Sets. Journal of Optimization Theory and Applications 126, 1–21 (2005)