Joint Virtual Reality Conference of EuroVR - EGVE (2011)
S. Coquillart, A. Steed, and G. Welch (Editors)

An interdisciplinary VR-architecture for 3D chatting with non-verbal communication

S. Gobron[†1], J. Ahn[1], Q. Silvestre[1], D. Thalmann[2], S. Rank[3], M. Skowron[3], G. Paltoglou[4], and M. Thelwall[4]

[1]IIG, EPFL, Switzerland; [2]NTU, Singapore; [3]OFAI, Austria; [4]Wolverhampton University, UK

Abstract

The communication between avatar and agent has already been treated from different but specialized perspectives. In contrast, this paper gives a balanced view of every key architectural aspect: from text analysis to computer graphics, the chatting system and the emotional model. Non-verbal communication, such as facial expression, gaze, or head orientation is crucial to simulate realistic behavior, but is still an aspect neglected in the simulation of virtual societies. In response, this paper aims to present the necessary modularity to allow virtual humans (VH) conversation with consistent facial expression -either between two users through their avatars, between an avatar and an agent, or even between an avatar and a Wizard of Oz. We believe such an approach is particularly suitable for the design and implementation of applications involving VHs interaction in virtual worlds. To this end, three key features are needed to design and implement this system entitled 3D-emoChatting. *First, a global architecture that combines components from several research fields. Second, a real-time analysis and management of emotions that allows interactive dialogues with non-verbal communication. Third, a model of a virtual emotional mind called* emoMind *that allows to simulate individual emotional characteristics. To conclude the paper, we briefly present the basic description of a user-test which is beyond the scope of the present paper.*

Categories and Subject Descriptors (according to ACM CCS): Three-Dimensional Graphics and Realism [I.3.7]: Virtual Reality—Natural Language Processing [I.2.7]: Text Analysis—

1. Introduction

Mainly due to entertainment industry requirements, virtual worlds such as landscapes, cities, and even solar systems, are becoming increasingly impressive in terms of technical effects. However, simulating human behavior –and especially realistic interactions between virtual humans (VH)– remains a challenging issue. A key type of interaction is inter-character non-verbal communication where emotions play an essential role. Social science and psychology has produced many interesting models concerning emotion, but it is also another challenge to apply them to interactive virtual environment. In the virtual reality (VR) domain, communication with a computer-driven VH (called an agent) is a well know research topic. Unfortunately, research including both semantic and emotional communication models is rare and always specialized. From our point of view, the main issues to design such a VR conversational system are the consistency and balance of its various aspects. It is not easy to understand well every required domain (CG, AI, data mining...). We believe it is even more difficult to integrate the corresponding engines together appropriately. Indeed, the system has to work constantly in real-time (constant 60 fps), which implies making design choices. For instance, graphically beautiful but computationally costly algorithms cannot be applied, VHs animation should be natural –implying complex issue relative to assembling MoCap–, lag between questions and responses must be reduced as much as possible but sometimes be lengthened to fake human delay variability. Our system answers these constraints, furthermore, each part can also be separately extended to allow maintenance and improvement.

In terms of applications, we believe that such a system would be particularly suitable for: (a) MMORPG games where agents would play the role of non-active characters such as merchant, bartenders, etc.; (b) immersive commercial sites in the context of specialized agents welcoming and answering FAQs to clients of a specific brand or product.

† (cor.) E-mail: stephane.gobron@epfl.ch

DOI: 10.2312/EGVE/JVRC11/087-094

The *affect bartender* shortly introduced as user-test in the last section is a direct application of our model.

In the following sections, we describe the overall architecture and process pipelines of each main interdisciplinary part of research.

2. Background

2.1. Virtual Reality

Emotional communication in virtual worlds has been a challenging research field over the last couple of decades. Cassell *et al.* [CPB*] proposed a system which automatically generates and animates conversations between multiple human-like agents with appropriate and synchronized speech, intonation, facial expressions, and hand gestures. Perlin and Goldberg [PG96] proposed an authoring tool (Improv) to create actors that respond to users and to each other in real-time, with personalities and moods consistent with the authors' goals and intentions. In this paper, we have not considered speech and hand gesture, however proposed the whole complex pipeline of VH conversation. Cassell *et al.* [CVB01] proposed a behavior expression animation toolkit (BEAT) that allows animators to input typed text to be spoken by an animated human figure. Compared to the BEAT system, the proposed framework mainly focus on visualization of emotional parameters extracted from chat sentence analysis. Su *et al.* [SPW07] predicted specific personality and emotional states from hierarchical fuzzy rules to facilitate personality and emotion control. Pelachaud [Pel09] developed a model of behavior expressivity using a set of six parameters that act as modulation of behavior animation. In our approach, we use 2D emotional parameters $\{v,a\}$ [RBFD03] that apply efficiently to conversations. In fields such as VR, computer vision, computer animation, robotics, and human computer interaction, efforts to synthesize or decode facial activity have recently been successful [CK07]. A *Facial Action Coding System (FACS)* [EF78] was developed that permits an objective description of facial movements based on an anatomical description. We derived our facial expression component from this FACS Action Units (AU).

2.2. Conversational Systems

Work in this area focuses on embodied conversational agents [PP01], and VH [GSC*08, KMT08]. Prominent examples of advances in this field are a framework to realize human-agent interactions while considering their affective dimension, and a study of when emotions enhance the general intelligent behavior of artificial agents resulting in more natural human-computer interactions [ABB*04]. Reithinger *et al.* [RGL*06] introduced an integrated, multi-modal expressive interaction system using a model of affective behavior, responsible for emotional reactions [BNP*05, STA*10] and presence of the created VH. Their conversational dialog engine is tailored to a specific, closed domain application, however: football-related game show settings. Similarly to our bartender application, Kopp *et al.* proposed in [KGKW05] a conversational agent as a museum guide. This study is complementary to the current paper as they focus on the dialog system and not on the general architecture.

2.3. Affective Linguistic Data Mining

Pang *et al.* [PLV02] were amongst the first to explore sentiment analysis, focusing on machine-learning approaches to analyze reviews [Seb02]. They experimented with three standard classifiers: Support Vector Machines (SVMs), Naive Bayes, and Maximum Entropy classifiers using a variety of features including simple words, phrases and structural elements. Mullen and Collier [MC04] used SVMs and enhanced the feature set with information from diverse sources. Dave *et al.* [DLP03] presented several feature extraction and scoring approaches for sentiment analysis, such as collocations, negation detection and substrings. They didn't report significant increases in comparison to other machine-learning approaches. Whitelaw *et al.* [WGA05] used fine-grained semantic distinctions in the feature set to improve classification. Their approach was based on a lexicon of adjectives with appraisal attribute values and modifiers. Zaidan *et al.* [ZEP07] experimented with so called *annotator rationales*: important words or phrases deemed significant for the polarity of reviews by human annotators. The disadvantage of their approach is that it requires additional human annotation on a specific data set, so it cannot be easily applied to open discussions and exchanges.

3. Overall Architecture

Human communication is first of all about social and psychological processes. Therefore, before presenting details of our communication pipeline, we first introduce the main motivation of our work: what are *meaning* and *emotion*?

3.1. Communication: Meaning and Emotion

Human communication is a multi-layered interactive system (dashed areas of Figure 1 outline factual and emotional layers) involving transactions between participants, relying not only on words, but also on a number of paralinguistic features such as facial/body expressions, voice changes, and intonation. These different channels provide rapidly changing contexts in which content and nonverbal features change their meaning. Typically in the interaction process not all information is successfully transmitted and receivers also perceive cues that are not really there [KHS91]. Situational, social and cultural contexts shape further what is being encoded and decoded by the interaction partners [Kap10]. Early research focused on reduced channel bandwidth in

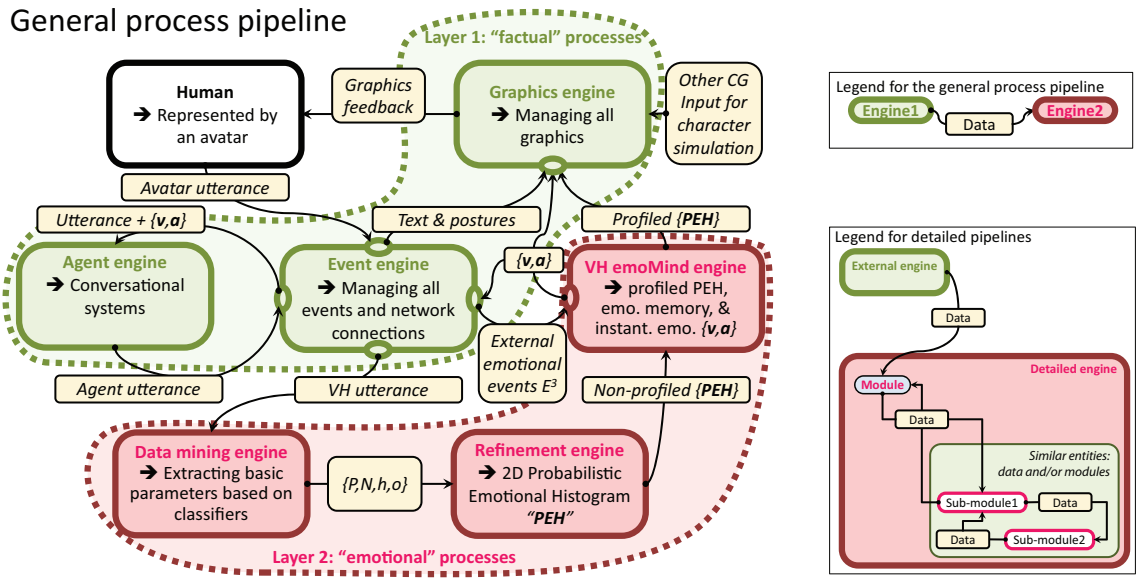

Figure 1: *Summary of the general process pipeline where the direct communication layer is represented by the three green engines and the non-verbal engines in red.*

text-only mediated communication and favored models emphasizing deficits. As mediated communication became increasingly multimodal and involved more realism in the representation of others, and particularly nonverbal communication, the situation became less clear (see " Virtual gestures: Embodiment and nonverbal behavior in computer-mediated communication" chapter in [KK11]. It is also quite possible that the addition of visual cues might not serve communication goals, despite being more interesting and pleasant (e.g., the chapter "Visual cues in computer-mediated communication: Sometimes less is more" in [KK11].

3.2. General process pipeline

Communication with a machine in a virtual world consists of at least: a user, a user-interface to manipulate an avatar, a dialogue/vocabulary analyzer, an emotional mind model, a graphics engine, and a listener framework playing the role of the agent. Most of the above, except the user, can be stored in two main layers: the factual processes layer, including the agent, the event, and the graphics engines; the emotion processes layer, including the data mining, refinement, and virtual human emoMind engines. Figure 1 presents the general structure of this process pipeline. Even if this architecture seems relatively complex, it remains a poor caricature of current knowledge.

All engines in this pipeline are described in Section 4.1. Figure 2 details the heart of all event management processes. Figure 3 illustrates the main concept behind the agent utterance questions and answers. Figure 4, associated with Figures 5 and 6, describes the emotion management component (from dictionaries for emotion analysis to emotion instantaneous states and memories). Figure 7 presents the 3D graphics management, focussing on VH facial emotion rendering and text chatting management. An avatar or an agent can start a conversation, and every utterance is a new event that enters the event engine and is stored in a queue. All communication is done via text.

Text utterances are analyzed by classifiers to extract potential emotional parameters, which are then refined to produce a *multidimensional probabilistic emotional histogram* (PEH) –see [GAP*10] and [BNP*05] for a practical example of multi-dimensional emotional model for interactive gaming with an agent. This generic PEH is then *personalized* depending on the character (*e.g.* for an optimist this would trend towards higher valence). Then the current *{v,a}*(valence and arousal emotional coordinate axes) state is combined with the PEH to yield the new state. The *{v,a}* values that result are transmitted to the interlocutor. If this is a conversational system (such as the *affect bartender* used in Section 5), then it produces a text response potentially influenced by emotion. In parallel, different animated postures are selected (*e.g.* idle, thinking, speaking motions). This process continues until the end of dialog.

4. Process Pipeline Engines

Our process pipeline consists of six modules (*engines*): event management; utterance response; emotion extraction; an "emoMind" system that profiles emotion depending on the virtual character's emotion refinement (two engines); and

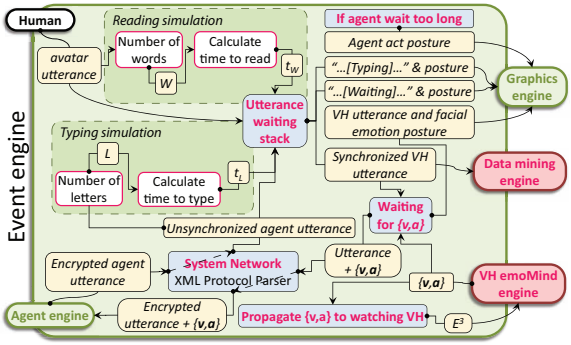

Figure 2: *Event engine, where the main tasks are to synchronize utterances, to manage graphical outputs (texts, facial expressions, and arms sequence of movements), and to simulate the interlocutor reading and typing so that agents and VH cannot be distinguished.*

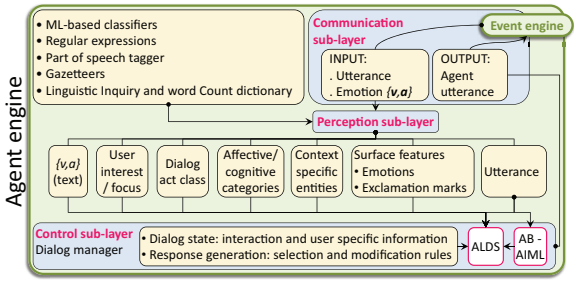

Figure 3: *The Agent engine with its three sub-layers (not to be confused with the main factual and emotional layers),* i.e. *communication, perception, and control.*

the animation and rendering engines. The video demonstration of the entire architecture can be found at:
http://3d-emo-chatting.serveftp.org/

4.1. Event engine

Our chat system is basically composed of two networks, one for [human (Avatar) $\rightleftharpoons$ *Wizard of Oz*(Woz)] interaction and one for [human (Avatar) $\rightleftharpoons$ machine (Agent)] interaction (the notion of "Woz" consists of a human pretending to be a machine). The user interface (UI) consists of two text windows. The top one displays the dialog for each session and the bottom one is the user input edit window. Details of the process pipeline for event management are illustrated in Figure 2. A key step to a consistent conversational system is the coherence of the emotional dynamic over time. Becker *et al.* [BW04] proposed an advanced study of this phenomenon. To reduce the complexity of our model, we developed a simple "Ping-Pong" effect to simplify the management of the dialogue between the virtual agent and the avatar. In the bar dialog, the bartender always starts the discussion usually by proposing something to drink just after a casual greetings. Each time the user type something, a thinking animation is executed by the agent. Then, when the user finalizes a sentence by pressing the "Enter" key, and after an artificial reading delay, the text utterance is sent to the Woz or Agent. If nothing happens during 25 seconds, the bartender simulates to clean the bar until a new event occurs.

4.1.1. Chatting with a conversation machine (CM)

We established a connection to communicate with a foreign server using a *XML-RPC* ("extensible markup language - remote procedure call") protocol. New utterances are sent to the conversational agent server when the corresponding emotion is computed, in parallel, the sentence is added to the dialog text box. We then wait for another event to occur. Working with long distance servers, we observed a delay of at most three seconds. To simulate the reading process, the message "...[Waiting]..." is shown during a delay proportional to the number of words. To simulate typing, when a message arrives, the message "...[Typing]..." is shown to the other dialog box for a duration propositional to the number of letters.

4.1.2. Chatting with the *Wizard of Oz*

In the case of Woz, he or she is hidden in a separate room. Similarly to the first protocol, "...[Typing]..." is shown when the Woz starts typing.

4.2. Agent engine

The conversational system produces natural language responses that will be played by the bartender as a VH in the virtual world and manages dialog between the virtual agent and users' avatar. The general system architecture consists of three sub-layers: communication, perception, and control. Figure 3 presents the Agent Engine architecture - used for the "Affect Bartender" [SPR*11] virtual agent - and its interface to the VR event engine.

4.2.1. Sub-layers

The **Communication sub-layer** provides an interface for reception and decoding of a user utterance and *{v,a}* values, which represent the emotional facial expression of the user's avatar. It also formats and dispatches system responses to the VR event engine. The **Perception sub-layer** integrates a number of natural language processing tools and affective states classifiers to analyze user utterances. In particular the "Affect Bartender" perception sub-layer includes: a maximum entropy based dialog act classifier, an utterance focus and utterance interest detector, regular expressions and gazetteers used for detecting instances of bar-context specific entities, a sentiment classifier [Section 3.4] and a *Linguistic Inquiry and Word Count* [PFB01] resource (*e.g.* for

assigning affective and cognitive categories). The **Control sub-layer** manages dialog with the user. It analyzes information obtained from the perception sub-layer, the observed dialog states, and information discovered in user utterances.

Two core components are applied for dialog management. The **Affect Bartender** ***Artificial Intelligence Markup Language*** **(AIML)** set (AB-AIML) [Wal01] provides a knowledge base specific to the virtual bartender and bar settings; responses for open-domain context, chats; and responses which convey the bartender's openness, interest in user feelings, current mood, recent events, etc. The **Affect Listener Dialog Scripting** (ALDS) as applied in the *Affect Bartender* condition is responsible for close-domain, task oriented dialog management (virtual bar, bartender context); providing responses based on affect-related states perceived during a dialog, either based on analysis of user utterances or *{v,a}* values attributed to the facial expressions of the user's avatar; and means to resolve situations where response candidates generated based on AB-AIML lack the necessary relevance to a given user utterance, *e.g.*, detection of "confusion statements" in responses returned from AB-AIML.

4.2.2. Role of affective cues in dialog management

Affective cues play an important role in response generation or modification of system response candidates, especially in situations when the system lacks information to generate a response based on other methods. Examples of mechanisms used when generating affective cue based responses include system comments and "affective feedback" related to: user utterance, based on detection of high positive or negative valence in a user utterances; significant changes detected based on *{v,a}* values representing an avatar's emotional facial expression, between two user utterances (*e.g.* "you look much more optimistic than just a second before... goooood :-) what happened?"); surface features used in user utterance (*e.g.* emoticons, usage of exclamation marks); affect, cognitive, social and linguistic categories discovered in user utterances (*e.g.* swear word category - "you look like a really decent person... please don't use this type of words excessively often ;-)"). The affective cues and affect-related dialog management mechanisms presented above enable the conversational system to generate responses that are not limited solely to semantic or pragmatic analysis of user utterances or a single pair of messages exchanged between the system and a user.

4.3. Emotional data mining engine

4.3.1. Emotion detection engine

We view the problem of detecting and extracting the emotions from text sentences as a classification problem. The general aims of classification is, given a document D and a pre-defined set of categories $C = \{c_1, c_2, ...c_t\}$, to assign D to one or more categories. We have approached the problem from two different perspectives, aimed at different classification sub-problems. The first one is an *unsupervised* lexicon-based classifier [GAP*10], which utilizes various emotionally-enhanced dictionaries to extract the emotional polarity and intensity of the textual input. The second is a *supervised*, machine-learning based meta-classifier which utilizes the output of three standard machine-learning classifiers in order to make a final prediction.

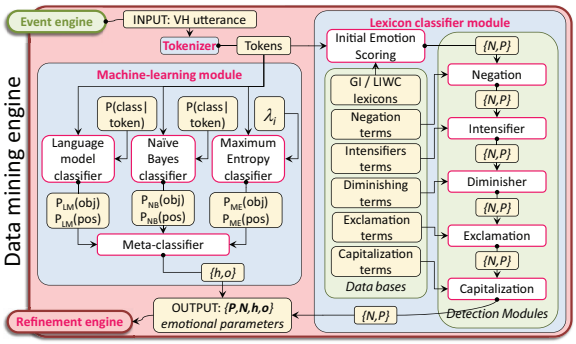

Figure 4: *This figure summarizes the data mining engine with its two main modules: the machine-learning module for text "happiness" and "objectivity" {*h,o*}, and the lexicon classifier module with* N,P *parameters.*

4.3.2. Lexicon-based classifier

The Lexicon-based classifier (right side of Figure 4) is based on two different emotional word-lists: The *General Inquirer* (GI) [SDSO66] and the *Linguistic Inquery and Word Count* (LIWC) [PFB01] lexicons. Those contain words with pre-assigned emotional indicators on a scale of $\{-5, ..., -1\}$ for negative terms and $\{+1, ..., +5\}$ for positive terms. The scales aim to capture the emotional intensity of each token. For example, in the latter lexicon, the word "love" is given an emotional weight of '+3' while "adore" has a weight of '+4'. The process of assigning specific scores at the tokens of the LIWC lexicon is described in detail in [TBP*10]. The GI dictionary provides only lists of positive and negative terms, so we simply assign a score of '+3' to all the positive and a score of '-3' to the negative. The Lexicon-based classifier scans the provided text and extracting the words that occur in either dictionary. Subsequently, the area around the extracted words is scanned for *emotion modifiers*, *i.e.* linguistically-driven features that change the polarity or intensity of the emotional words. Those include: negators, intensifiers, diminishers, emoticons, exclamations, and fully-capitalized words. The classifier's output is two scores, one for the positive $\{+1, ..., +5\}$ and one for the negative $\{-5, ..., -1\}$ dimension, defined as P ("Positive") and N ("Negative").

4.3.3. Machine-learning meta-classifier

The machine-learning meta-classifier (left side of Figure 4) uses as input the output of three individual machine-learning classifiers to make a final estimation. Specifically, we use three standard, probabilistic, state-of-the-art classifiers: a *Language Model* [PSW03], a *Naive Bayes* [MS99] and a *Maximum Entropy* [NLM99] classifier. All classifiers function in a two-tier fashion. The first-stage classification determines the probabilities of whether D is objective or subjective ($C_1 = \{obj, sub\}$) and the second-stage classification determines the probabilities of the polarity of the document ($C_2 = \{neg, pos\}$). A document is considered subjective if it contains expressions of opinion, emotion, evaluation, speculation etc, overall defined as *private states* [QGLS85]. It is considered objective if it contains factual information and there are no expressions of private states. Additionally, a document is considered positive if it contains positive expressions of opinion, emotion or evaluation and negative if it contains negative expressions. Therefore, for a document D the outcome of the classifiers is $\{P_x(obj|D), P_x(pos|D)\}$ where $x = \{LM, NB, MaxEnt\}$ for each of the classifiers used respectively. The events $\{neg, pos\}$ are complimentary, therefore $P(neg|D) = 1 - P(pos|D)$. The same is also true for the events $\{obj, sub\}$. The produced probabilities are provided to the meta-classifier which averages their value and produces a final output for *o*bjectivity: $o = P_{meta}(obj|D) = \frac{1}{|x|}\sum_x P_x(obj|D)$ and *h*appiness: $h = P_{meta}(pos|D) = \frac{1}{|x|}\sum_x P_x(pos|D)$. The meta-classifier's purpose is to moderate the effects of any individual first-level classifier: in the event that any one of them produces biased results, the final output isn't similarly biased as it will have been moderated by the results of the other two classifiers.

4.3.4. Training

We trained the machine-learning classifiers on the BLOG dataset [MOS08]. The dataset is comprised of an uncompressed 148 GB crawl of approximately 100,000 blogs and their respective RSS feeds. The dataset has been used for 3 consecutive years by the Text REtrieval Conferences. Participants of the conference are provided with the task of finding documents (*i.e.* blog posts) expressing an opinion about specific entities X, which may be people, companies, films etc. The results are given to human assessors who then judge the content of the posts and assign each one a score: for instance, "1" if the document contains relevant, factual information about the entity but no expression of opinion and "2" if the document contains an explicit negative opinion towards the entity. We used the assessments from all 3 years of the conference to train our classifiers, resulting in 200 different entity searches and 11,137 documents. For the second stage classifier (*i.e.* $C_2 = \{pos, neg\}$), we used the documents assigned a "2" as negative and "4" as positive.

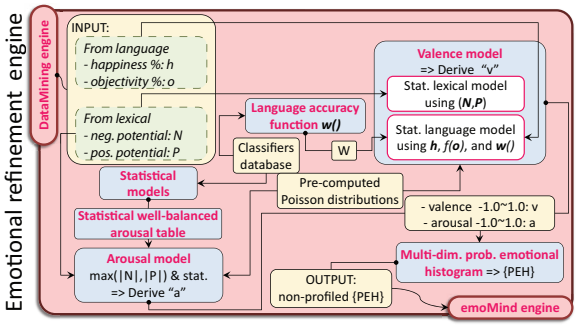

Figure 5: *The refinement engine mainly transforms the {*h,o,N,P*} values extracted from the data mining engine into a 2D table of potentials - representing valence and arousal {*v,a*} - called* PEH*; details can be found in [GAP*10].*

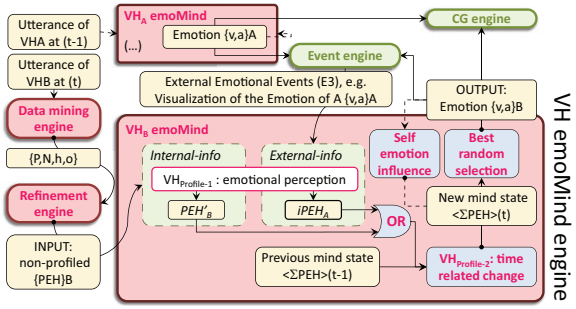

Figure 6: *The emotional mind engine is the center of VH affect where events are profiled depending on the predefined mind status, memory of emotional events is kept, and where instantaneous emotions are selected.*

4.4. VH emoMind engine

Four parameters (N, P, h, and o) with different properties (*i.e.* range, intensities, means) are extracted from the text using the classifier of the data mining engine, all of them influencing in different ways the emotion we would like to deduce. There is no universal well defined model of emotion, and this paper does not pretend to solve the complexity of emotion. Nevertheless, we tried to establish a model where emotion can be interpreted, profiled, and stored for simulating memories of past emotion - similarly to *a state of mind*. For this, we designed a PEH that uses as input parameters extracted from the text utterance. The data chart of the emotional refinement is in Figure 5, and a PEH is also illustrated in the upper area above the VH head of Figure 8(d), including VH personality settings and emotional memory affects.

The purpose of the *emotional mind* (emoMind) engine is to influence a non-profiled, generic PEH with the virtual character's mind characteristics. As an example, in the user-test (Section 5), we set the *affect bartender* to have a dynamic and optimist profile with minimum affective mem-

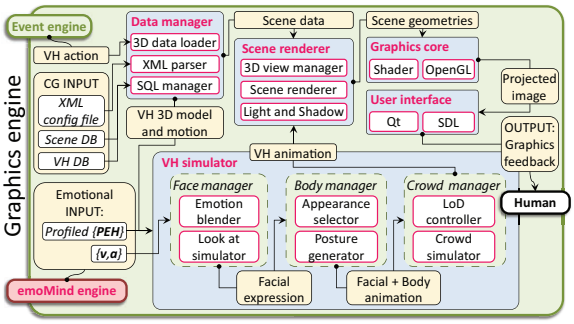

Figure 7: *Graphics engine is similar to classical crowd engines with the specificity of the emotional events.*

ory and no self-emotion influence threshold. His emotional state was strongly responsive to positive events and would decrease the effect of low arousal or negative valence values. Figure 6 shows the main functions of the emoMind engine. The *historical emotional event storage* is illustrated in this figure where new emotional event and previous mind state produce new mind states. Graphical interpretation of *emotional perception* can also be found in the lower area above the agent head of Figure 8(d).

4.5. Computer graphics engine

As depicted in Figure 7, our graphics engine simulates the VH's facial expression and body movement by getting data from the "Data manager" and sending all the simulated animation to the "Scene renderer". The VH actions and *{v,a}* values from the Event and emoMind engines are inputs to our graphics engine. For the facial expression, we analyzed min-max rotational angles. The relations between *{v,a}* values and moving facial parts are based on FACS AU [EF78] [GAP*10]. The *{v,a}* parameter from the emoMind engine controls these facial joint angles for emotional expression. The "Facial manager" controls emotional expression and gaze detection with 14 degrees of freedom. The proposed event engine triggers actions such as body movements (stand, idle, think, talk, listen, look, work, walk, run, and sit). In the user-test, the state of body motion were transferred from condition by condition. For some actions, we used one of several captured motions, chosen at random each time, to improve the realism of the animated scene.

5. Conclusion and User-test

In this paper, we have presented a VR architecture entitled *3D-emoChatting*, enabling chatting dialog with semantic (*i.e.* text utterances) and induced emotional communication based on valence and arousal emotional dimensions. Since, all interdisciplinary aspects of verbal and non-verbal communications are included (*i.e.* data mining, VR, CG, distant protocols, artificial intelligence, and psychology), we believe this model to be of a real practical use for entertainment applications involving virtual societies.

To validate our model, a user-test involving 40 participants was performed with four experiments: with or without facial emotion and with a conversational system or a *Wizard of Oz*. The virtual scenes of our user-test are depicted in Figure 8. However, the presentation, structure, and questionnaire analysis of this user-test are beyond the scope of this paper and is currently the object of another submission.

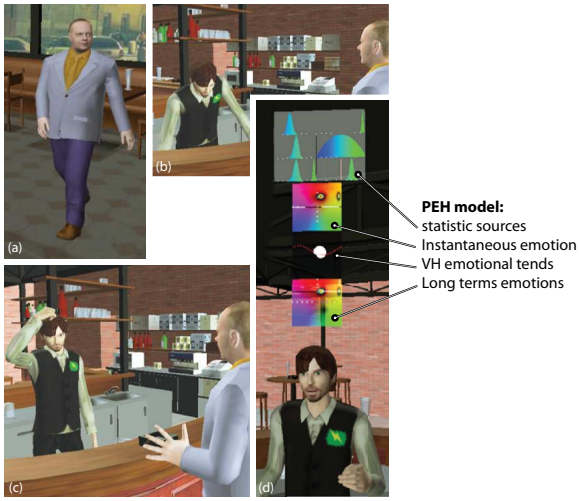

Figure 8: *Main steps during the user-test resulting from the CG engine: a) the avatar goes into the bar; b) when no interaction, the agent simulates cleaning the bar; c) a "Can I have a <any drink>?", the agent simulates to think before answering; d) The agent answers and serves with a facial expression (notice the PEH above his head for testing).*

Acknowledgements

This research has been funded by a European Union grant, 7th Framework Programme, Theme 3: *Science of complex systems for socially intelligent ICT*, which is part of the CYBER**EMOTIONS** Project (Contract 231323).

References

[ABB*04] ARAFA Y., BOTELHO L., BULLOCK A., FIGUEIREDO P., GEBHARD P., HOOK K., MAMDANI E., PAIVA A., PETTA P., SENGERS P., VALA M.: Affective interactions for real-time applications: the safira project. *KI-Journal 18*, 30 (2004). 2

[BNP*05] BECKER C., NAKASONE A., PRENDINGER H., ISHIZUKA M., WACHSMUTH I.: Physiologically interactive gaming with the 3d agent max. In *Intl. Workshop on Conversational Informatics* (2005), pp. 37–42. 2, 3

[BW04] BECKER C., WACHSMUTH I.: Simulating the emotion dynamics of a multimodal conversational agent. In *In Proceedings Tutorial and Research Workshop on Affective Dialogue Systems (ADS-04), LNAI 3068* (2004), Springer, pp. 154–165. 4

[CK07] COHN J. F., KANADE T.: Automated facial image analysis for measurement of emotion expression. In *The handbook of emotion elicitation and assessment* (2007), Oxford University Press Series in Affective Science, pp. 222–238. 2

[CPB*] CASSELL J., PELACHAUD C., BADLER N., STEEDMAN M., ACHORN B., BECKET T., DOUVILLE B., PREVOST S., STONE M.: Animated conversation: rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents. In *SIGGRAPH '94*. 2

[CVB01] CASSELL J., VILHJÁLMSSON H. H., BICKMORE T.: Beat: Behavior expression animation toolkit. In *SIGGRAPH'01* (2001), pp. 477–486. 2

[DLP03] DAVE K., LAWRENCE S., PENNOCK D. M.: Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In *Proceedings of the 20th international conference on World Wide Web* (2003), pp. 519–528. 2

[EF78] EKMAN P., FRIESEN W.: Facial action coding system. *Consulting Psychologists Press* (1978). 2, 7

[GAP*10] GOBRON S., AHN J., PALTOGLOU G., THELWALL M., THALMANN D.: From sentence to emotion: a real-time three-dimensional graphics metaphor of emotions extracted from text. *The Visual Computer 26*, 6-8 (2010), 505–519. 3, 5, 6, 7

[GSC*08] GEBHARD P., SCHROEDER M., CHARFUELAN M., ENDRES C., KIPP M., PAMMI S., M. R., O. T.: Ideas4games: Building expressive virtual characters for computer games. In *In Proceedings of the 8th International Conference on Intelligent Virtual Agents* (2008), LNAI, Springer, pp. 426–440. 2

[Kap10] KAPPAS A.: Smile when you read this, whether you like it or not: Conceptual challenges to affect detection. *IEEE Transactions on Affective Computing 1*, 1 (2010), 38–41. 2

[KGKW05] KOPP S., GESELLENSETTER L., KRÄMER N. C., WACHSMUTH I.: A conversational agent as museum guide - design and evaluation of a real-world application. In *The 5th International Working Conference on Intelligent Virtual Agents (IVA'05* (2005), Springer, pp. 329–343. 2

[KHS91] KAPPAS A., HESS U., SCHERER K. R.: Voice and emotion. In *Fundamentals of nonverbal behavior* (1991), Cambridge University Press, pp. 200Ű–238. 2

[KK11] KAPPAS A., KRÄMER N.: *Face-to-face communication over the Internet*. Cambridge: Cambridge Univ. Press, 2011. 3

[KMT08] KASAP Z., MAGNENAT-THALMANN N.: Intelligent virtual humans with autonomy and personality: State-of-the-art. In *New Advances in Virtual Humans*, vol. 140. Springer Berlin / Heidelberg, 2008, pp. 43–84. 2

[MC04] MULLEN T., COLLIER N.: Sentiment analysis using support vector machines with diverse information sources. In *Proceedings of EMNLP 2004* (Barcelona, Spain, July 2004), pp. 412–418. 2

[MOS08] MACDONALD C., OUNIS I., SOBOROFF I.: Overview of the trec-2008 blog track. In *The Sixteenth Text REtrieval Conference (TREC 2008) Proceedings* (2008). 6

[MS99] MANNING C. D., SCHUETZE H.: *Foundations of Statistical Natural Language Processing*. The MIT Press, 1999. 6

[NLM99] NIGAM K., LAFFERTY J., MCCALLUM A.: Using maximum entropy for text classification. In *IJCAI-99 Machine Learning for Information Filtering* (1999), pp. 61–67. 6

[Pel09] PELACHAUD C.: Studies on gesture expressivity for a virtual agent. *Speech Commun. 51*, 7 (2009), 630–639. 2

[PFB01] PENNEBAKER J., FRANCIS M., BOOTH R.: *Linguistic Inquiry and Word Count*, 2 ed. Erlbaum Publishers, 2001. 4, 5

[PG96] PERLIN K., GOLDBERG A.: Improv: a system for scripting interactive actors in virtual worlds. In *SIGGRAPH'96* (1996), ACM, pp. 205–216. 2

[PLV02] PANG B., LEE L., VAITHYANATHAN S.: Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (2002). 2

[PP01] PELACHAUD C., POGGI I.: Towards believable interactive embodied agents. In *Fifth Int. Conf. on Autonomous Agents workshop on Multimodal Communication and Context in Embodied Agents* (2001). 2

[PSW03] PENG F., SCHUURMANS D., WANG S.: Language and task independent text categorization with simple language models. NAACL '03, Association for Computational Linguistics, pp. 110–117. 6

[QGLS85] QUIRK R., GREENBAUM S., LEECH G., SVARTVIK J.: *A Comprehensive Grammar of the English Language*. Longman, 1985. 6

[RBFD03] RUSSELL J. A., BACHOROWSKI J.-A., FERNANDEZ-DOLS J.-M.: Facial and vocal expressions of emotion. In *Annual Review of Psychology* (2003), pp. 329–349. 2

[RGL*06] REITHINGER N., GEBHARD P., LOECKELT M., NDIAYE A., PFLEGER N., KLESEN M.: Virtualhuman - dialogic and affective interaction with virtual characters. In *In Proceedings of the 8th International Conference on Multimodal Interfaces* (2006). 2

[SDSO66] STONE P. J., DUNPHY D. C., SMITH M. S., OGILVIE D. M.: *The General Inquirer: A Computer Approach to Content Analysis*. MIT Press, 1966. 5

[Seb02] SEBASTIANI F.: Machine learning in automated text categorization. *ACM Computing Surveys 34*, 1 (2002), 1–47. 2

[SPR*11] SKOWRON M., PIRKER H., RANK S., PALTOGLOU G., AHN J., GOBRON S.: No peanuts! affective cues for the virtual bartender. In *Proc. of the Florida Artificial Intelligence Research Society Conf.* (2011), AAAI Press, pp. 117–122. 4

[SPW07] SU W.-P., PHAM B., WARDHANI A.: Personality and emotion-based high-level control of affective story characters. *IEEE Transactions on Visualization and Computer Graphics 13*, 2 (2007), 281–293. 2

[STA*10] SWARTOUT W., TRAUM D., ARTSTEIN R., NOREN D., DEBEVEC P., BRONNENKANT K., WILLIAMS J., LEUSKI A., NARAYANAN S. S., PIEPOL D., LANE C., MORIE J., AGGARWAL P., LIEWER M., CHIANG J.-Y., GERTEN J., CHU S., WHITE K.: Ada and grace: Toward realistic and engaging virtual museum guides. In *In Proceedings of the 10th International Conference on Intelligent Virtual Agents (IVA)* (September 2010). 2

[TBP*10] THELWALL M., BUCKLEY K., PALTOGLOU G., CAI D., KAPPAS A.: Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology 61*, 12 (2010), 2544–2558. 5

[Wal01] WALLACE R.: Don't read me - a.l.i.c.e. and aiml documentation. ACM SIGGRAPH 2002 Course #16 Notes, http://www.alicebot.com/dont.html 2001. 5

[WGA05] WHITELAW C., GARG N., ARGAMON S.: Using appraisal groups for sentiment analysis. In *CIKM '05: Proceedings of the 14th ACM international conference on Information and knowledge management* (New York, NY, USA, 2005), ACM, pp. 625–631. 2

[ZEP07] ZAIDAN O., EISNER J., PIATKO C.: Using annotator rationales to improve machine learning for text categorization. *NAACL HLT* (2007), 260–267. 2