



**Österreichisches Forschungsinstitut für /  
Austrian Research Institute for /  
Artificial Intelligence**

**TR-2011-02**

*Marcin Skowron, Georgios Paltoglou*

**Affect Bartender - Affective Cues and Their  
Application in a Conversational Agent**

- Freyung 6/6 • A-1010 Vienna • Austria •
- Phone: +43-1-5336112 •
- <mailto:sec@ofai.at> •
- <http://www.ofai.at/> •



**Österreichisches Forschungsinstitut für /  
Austrian Research Institute for /  
Artificial Intelligence**

**TR-2011-02**

*Marcin Skowron, Georgios Paltoglou*

**Affect Bartender - Affective Cues and Their  
Application in a Conversational Agent**

The Austrian Research Institute for Artificial Intelligence is supported by the  
Austrian Federal Ministry for Science and Research and the  
Austrian Federal Ministry for Transport, Innovation and Technology.

---

*Citation:* Skowron M., Paltoglou G.: Affect Bartender - Affective Cues and Their Application in a Conversational Agent. Technical Report, Österreichisches Forschungsinstitut für Artificial Intelligence, Wien, TR-2011-02, 2011

# Affect Bartender - Affective Cues and Their Application in a Conversational Agent

Marcin Skowron,  
Austrian Research Institute for Artificial Intelligence  
Vienna, Austria  
marcin.skowron@ofai.at

Georgios Paltoglou,  
University of Wolverhampton  
Wolverhampton, UK  
g.paltoglou@wlv.ac.uk

**Abstract**—This paper presents methods for the detection of textual expressions of users’ affective states and explores an application of these *affective cues* in a conversational system – Affect Bartender. We also describe the architecture of the system, core system components and a range of developed communication interfaces. The application of the described methods is illustrated with examples of dialogs conducted with experiment participants in a Virtual Reality setting.

**Index Terms**—Affective Interactions, Conversational Agent, Textual Affect Sensing, Sentiment Classification

## I. INTRODUCTION

Interactive environments that host business, educational or entertainment applications, e.g., online customer service centers, virtual training systems or immersive virtual worlds, benefit from the introduction of realistic autonomous characters such as online representatives and consultants, virtual tutors or non-playable characters. Such artificial entities serve tasks that often cannot be handled by human operators in a sustainable fashion because of restrictions on the availability of qualified staff, scalability, responsiveness, high costs, etc.

One important aspect of such applications is to deal with the affective states of their participants. In many application domains, for virtual autonomous characters to successfully conduct their tasks in the long term requires a system to be competent in both exchanging task-specific information with the users as well as establishing, maintaining and developing a personalized connection with them. I.e., apart from the goal-oriented information exchange systems have to also take other aspects of the ongoing communication into account, e.g., a user’s affective state or his/her satisfaction level. It requires systems capable of detecting changes in a user’s affective states and suitably reacting to them by incorporating emotional cues to the generated autonomous characters behaviour. For the purpose of this work, we define *affective cues* as indicative evidence of a user’s affective state that can be perceived by the agent; in particular, in the case of a conversation system, these relate to the textual expressions of user’s affective states.

Sentiment Analysis (also known as Opinion Mining) is the area of research that deals with the computational treatment of expressions of private states (i.e. personal states that are not open to objective observation or verification) in written speech. Its aim is to extract, assess and understand the nature and general properties of opinions, sentiments and emotions

explicitly expressed or implied in textual communication. Therefore, as human users are able to assess the emotional state of their conversational partner and accordingly adapt the communication style of any task-specific information to suit the particular situation, sentiment analysis provides this important function to conversational agents, with the overarching goal of approximating the human response as faithfully as possible.

This paper focuses on the description of a conversational agent – *Affect Bartender* and on the presentation on how the information on textual expressions of user’s affective states is applied in the dialog management. After an overview of relevant works from the areas of textual affect sensing and affective conversational agents, the architecture of the agent is presented. This section also describes the system layers, the core software components and introduces the system communication interfaces. In section 4, we focus on the methods applied for the analysis of textual expressions of users’ affective states. Next, we describe how the acquired information is incorporated into the dialog management component of the conversational agent and present excerpts of dialogs between the Affect Bartender and an experiment participant.

## II. RELEVANT RESEARCH

Opinion mining has been a popular research topic within the natural language processing (NLP) community, because it presents both interesting academic questions and significant commercial interest. The latter issue in particular has directed most of the research towards the analysis of product reviews [1], [2], where the overall aim is either to predict the number of stars given to a product by the reviewer or in the binary case to correctly identify whether the author recommends the product, based on the textual content of the review. Nonetheless, research into other kinds of textual content, such as political debates [3], news [4] and blogs [5] has also been presented recently. In this paper, we are interested in the identification of the affective states of users in a conversational system. The textual content of such systems usually comprises of short, informal-style textual exchanges, which is significantly different from the content of reviews, which tend to be more lengthy and syntactically and orthographically correct, as they are usually addressed to a much wider audience. Our approach

is therefore mostly relevant to the work of Neviarouskaya et al. [6] which try to predict the emotion of users based on the psychological theory of *basic emotions* and Thelwall et al. [7], which focuses on short, informal comments posted on MySpace and provides a ordinal prediction of valence in both the positive and negative dimension.

The design and development of artificial systems that integrate the modelling of emotional behavior focuses an interest of the research community[8][9][10]. Work in this field includes the recognition of affective states in human-computer interaction, in particular the analysis of multi-modal inputs: auditory modality [11], [12], [13] visual modality (e.g, facial expressions)[14], as well as their combination [15] [16] [17]. The research findings are applied, among others, in interactive systems, including e.g., the embodied conversational agents (ECAs), and Virtual Humans (VH) [18], [19] [20] [21].

The management of human-computer conversations that consider emotional cues is also a central area of interest in the design of Affective Dialog Systems (ADSs). This multidisciplinary field accommodates contributions from a range of research areas, e.g., speech recognition, dialog processing, computer graphics, animation, speech synthesis, embodied conversational agents and human-computer interaction [22], [23]. Numerous works provide significant evidence that emotional factors play a crucial role in human-computer interaction systems. For example, in the development of ECAs the integration of affective components enhances agents believability, tutoring systems that take into consideration the motivational states of students and attempt to suitably influence them can support the learning process [24].

### III. CONVERSATIONAL AGENT – SYSTEM ARCHITECTURE

The *Affect Bartender*, i.e., the conversational agent presented in this work is responsible for the management of the verbal communication between a VH (the virtual bartender) and a user, represented in the virtual 3D bar by an avatar (virtual client). The motivation for the incorporation of affective cues in the Affect Bartender’s perception and dialog management layers follows the empirical findings of Reeves and Nass [25]. There, the authors reported that users apply social norms to computers and that the user’s satisfaction can be enhanced when interacting with systems that are capable to perceive the user’s affective cues and to suitably incorporate them in their behaviour.

The main objectives for the system in the above presented interaction scenario included:

- 1) achieving realistic dialogs,
- 2) providing an enjoyable overall chatting experience,
- 3) establishing and maintaining emotional connection with the users.

The implementation of the system is based on the concept of *Affect Listeners* [26], conversational agents aiming to detect and adapt to affective states of users (i.e., textual expression of users affective states), and meaningfully respond to users’ utterances both at the content- and affect-related level.

The core tasks of the Affect Bartender in the context of the virtual bar scenario include:

- perception and classification of affective cues from user utterances and system response candidates,
- incorporation of the affective cues (based on text-analysis or provided by the 3D VR event engine that represent the emotional facial expression of the user’s avatar) into the dialog management,
- developing an emotional connection with the users (affective dialog management),
- management of task-oriented dialogs (closed-domain dialog),
- ensuring robust communication capabilities for conversations not restricted in topic (open-domain chats),
- detection of cues in the system-user interactions that enable the selection of suitable system response generation method (task oriented dialog vs. open-domain conversations).

At the top level, the system architecture consists of 3 layers: communication, perception and control. Figure 1 presents the layers of the system architecture and the interaction loop with the environment.

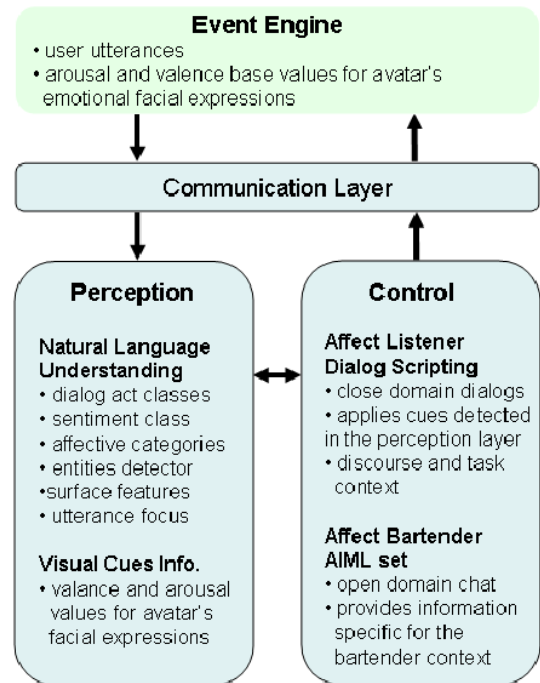


Fig. 1. Layers of the conversational system architecture and the interaction loop with the environment.

#### A. Perception Layer

The Perception Layer integrates a number of natural language processing tools, linguistic and affective resources to analyze the user utterances and system response candidates. The presented implementation of the Affect Bartender’s Perception Layer include in particular the following components:

- Dialog Act (DA) classifier – dialog act classes are based on the annotation schema used in the NPS Chat Corpus [27]. Based on the requirements of the Affect Bartender scenario, we decided to extend the originally used taxonomy of DA classes (Accept, Bye, Clarify, Continuer, Emotion, Emphasis, Greet, No Answer, Other, Reject, Statement, Wh-Question, Yes Answer, Yes/No Question), with an additional class “Order” (i.e. for ordering drinks). For this additional class 339 training instances were provided. The original NPS Chat class “System”, irrelevant for the system-user dialogs, was excluded along with the set of corresponding training instances. For the presented taxonomy and training set, the Maximum Entropy [28] based DA classifier<sup>1</sup> achieved 10-fold cross validation accuracy of 71.2%,
- Linguistic Inquiry and Word Count - LIWC (LC). This lexical resource provides a classification of words along 64 linguistic, cognitive, and affective categories [29], cf. section IV-A,
- Sentiment Classifier (SC) – provides information on: sentiment class (SC) i.e., negative {-1}, neutral {0}, and positive {1}. Further it assigns positive sentiment value (PS) {+1, ..., +5} and negative sentiment value (NS) {-5, ..., -1} of user utterances and system response candidates, cf. section IV-B,
- Regular expressions and gazetteers used for detecting instances of bar-context specific entities, i.e., drinks (DR) and snacks (SN),
- Surface features detector – e.g., exclamation marks (EM), emoticons (e.g., EE-sad, EE-smile),
- Utterance focus (UF) and utterance interest (UI) detector [30].

Figure 2 presents the excerpt of the Perception Layer annotation for example inputs: “I really like being here! :=)”, “What else can you offer to drink?”, “Do you always have so many customers?”, “You must be kidding! No more pina colada?”, and “Didn’t you forget to pay the bill last time??”.

### B. Control Layer

The layer manages the progression of a dialog by relating the observed dialog states to the intended ones (e.g., querying and follow-up questions on the user’s affective states, conducting specific bartender tasks). In this process, the Control Layer incorporates cues acquired by the Perception Layer described above (e.g. linguistic or affective categories discovered in a user utterance). The Control Layer generates a number of response candidates and then selects the system response from these. In this processes, the component integrates the rule-based action selection – Affect Listener Dialog Scripting (ALDS) with the command interpreter for the Affect-Bartender AIML-set. Both are described in sections 5.

<sup>1</sup>MaxEnt - [http://homepages.inf.ed.ac.uk/lzhang10/maxent\\_toolkit.html](http://homepages.inf.ed.ac.uk/lzhang10/maxent_toolkit.html) - Maximum Entropy Modeling Toolkit.

Input	Perception Layer Output (excerpt)
I really like being here! :=)	DA-Statement SC-1 NS- -1 PS-5 LC-Affect:Posemo :CogMech DR-0 EE-smile
What else can you offer to drink?	DA-whQuestion SC-0 NS- -1 PS-3 DR-0 LC-Social:Ingest :Leisure
Do you always have so many customers?	DA-ynQuestion SC- -1 NS- -1 PS-1 DR-0 LC-Social:Quant :Work UF-cutomer
You must be kidding! No more pina colada?	DA-Statement NS- -3 PS-1 LC-Social:Discrep:Certain DR-pina colada EM-1
Didn’t you forget to pay the bill last time??	DA-ynQuestion Sc- -1 NS-3 PS-3 LC-Money:Work:Past :Negate

Fig. 2. Perception Layer – annotation example.

### C. Communication Layer

The Communication Layer enables the conversational system to connect to a range of communication interfaces. This includes channels that are predominantly textual, e.g., Jabber<sup>2</sup>, ICQ<sup>3</sup>, Google Talk<sup>4</sup>, Facebook Chat<sup>5</sup>, IRC<sup>6</sup> and a webchat interface (web browser based). Further, the Communication Layer can also mediate connection between the Affect Bartender system and a Virtual Reality Event Engine that integrates the conversational system with a Virtual Human, e.g., Virtual Bartender [31]. In the later case, the Affect Bartender and 3D VR Event Engine can be situated on separate remote servers and the Communication Layer can handle their connection via the XMLRPC protocol<sup>7</sup>. In particular, in the settings that incorporate textual and visual cues, the Communication Layer can receive and decode user utterances as well as values on arousal and valence which are calculated as part of the process for generating emotional facial expressions of avatars and Virtual Humans. Further, the layer formats and dispatches the system responses.

## IV. AFFECTIVE CUES – PERCEPTION AND CLASSIFICATION

The ability to detect and to classify textual expressions of affective states in utterances of the users is a core prerequisite for the Affect Listener/Affect Bartender systems. In the described realization of the conversational agent two affect detection and classification methods are used: Sentiment Classifier [32] and Linguistic Inquiry and Word Count (LIWC) dictionary [33]. These resources provide information on the affective cues in user utterances or in system response candidates.

<sup>2</sup><http://xmpp.org/>

<sup>3</sup><http://www.icq.com/>

<sup>4</sup><http://www.google.com/talk/>

<sup>5</sup><http://www.facebook.com/sitetour/chat.php>

<sup>6</sup><http://www.irchelp.org/>

<sup>7</sup><http://www.xmlrpc.com/>

### A. Linguistic Inquiry and Word Count

The Linguistic Inquiry and Word Count (LIWC) dictionary used in the Affect Bartender's Perception Layer, enables the conversational agent to detect 64 linguistic, cognitive and emotional dimensions. The applied, regular expressions based, search mechanism discovers the matches between the word in the inputted text snippets (i.e., user utterance, system response candidates) and the entries from LIWC dictionary.

Among others, this resource provides 32 word categories that are tapping psychological processes (e.g., affective such as positive and negative emotions; social such as family, friends and human; cognitive such as insight, causation, tentative), 22 linguistic categories (e.g., adverbs, negations, swear words), 7 personal concern categories (e.g., home, religion, work, leisure) 3 paralinguistic dimensions (fillers, assents, nonfluencies), for almost 4500 words and word stems [33]. For example, the word "compassion" is categorized in 3 categories: affective processes, positive emotion and social processes; the word "grief" in 4 categories: affective processes, negative emotion, sadness (psychological processes) and death (personal concern). The dictionary also uses word stems such as e.g., "drunk\*", that enables the classification of words like "drunken" and "drunkenness" with categories like: biological processes, ingestion and leisure. In the recent years, LIWC has been successfully applied in various psychological and psycholinguistic studies that included e.g., the investigation of linguistic style, the relations between language use and speakers personality [34].

The system's Perception Layer applies the LIWC dictionary to detect words in the user utterances and system response candidates which are related to affective categories such as positive emotion, negative emotion, anger, sadness and anxiety. Further, the dictionary provides also cues about other categories useful for managing the system user communication, taking into consideration the agent's tasks i.e., expressions from categories such as: health, swear words, leisure, or work.

### B. Sentiment Classifier

Previous research has shown that although machine-learning techniques are very effective in the domain of product reviews, simpler, lexicon-based approaches are more appropriate when analysing short, informal exchanges [32], [7], [6] and therefore we adopt the latter approach. Lexicon-based approaches are typically based on manually or semi-automatically constructed lists of words, each annotated with a score, indicating their emotional content. In our experiments we used two different emotional word-lists: The "General Inquirer" and "Linguistic Inquiry and Word Count" (LIWC) software<sup>8</sup>, the latter as enriched by [7]. We utilize the emotional indicators assigned to the words in those dictionaries and enhance the final prediction by taking into consideration simple, linguistically-driven signals, such as negation detection, capitalization, intensifier/diminisher identification, emoticon detection, etc., which contribute in modifying the final emotional weight assigned to

a phrase. For example the word "love" has an initial score of +4 in the latter dictionary, but if detected in a sentence with exclamation marks its score increased by 1, to a final score of +5. Respectively if a negation is detected before the word, then its valence is reversed and decreased by one to a final score of -3. The algorithm outputs a vector of two scores, one for the positive scale and one for the negative on the format {neg,pos}, where neg={-5,...,-1} and pos={+1,...,+5}. Higher absolute values indicate higher emotional content in that dimension and {-1,+1} indicate lack of emotion. For example {-4,+1} would indicate a strong negative emotion, {-1,+5} a very strong positive emotion and {-3,+4} a mixed emotional response where the positive emotion slightly overwhelms the negative. The emotional score of a textual utterance is estimated as the maximum score in both the positive and negative dimensions. For example, the phrase "I hate the fact that I love you so much!!" would have a final score of {-3,+4} because of the existence of both negative words (i.e., "hate") and positive words (i.e., "love"), the score of the latter increased due of the presence of exclamations marks and the intensifier expression so much. More information about the classifier as well as a comparison of its effectiveness against state-of-the-art machine learning classifiers can be found in [32].

## V. AFFECTIVE CUES – APPLICATION IN THE DIALOG MANAGEMENT

Affective cues play an important role in the generation of response candidates as well as response modification and selection processes.

In particular, the agent uses information on the affective states when one of the following conditions are met:

- detection of a high arousal in a user utterance (Sentiment Classifier),
- rapid change in the user's affective states (based on two consecutive message exchange turns, i.e., analysis of user utterances or {v,a} values of the avatar's emotional facial expressions),
- recognition of a system response that was generated by a 'confusion statement' template, (i.e., the system was unable to retrieve a response that has a direct relevance to the factual content of a user's utterance), paired with the detection of a particular affective category in an utterance of the user (LIWC).

In the first situation the detection of a high arousal is based on the information acquired from the Sentiment Classifier presented in section IV-B. In the presented system, the activation threshold was set to +5 for the positive and -5 for the negative sentiment. In the practical interaction settings, these values signalize the usage of highly emotional words, which are often found in the utterances that contain information important for the user or which convey strongly emotional expressions.

The recognition of a rapid change of the observed affective states, as expressed by the emotional facial expressions of the user's avatar, enables the dialog manager component to relate to the visual cues that are perceived by the user during the interaction. This affective cue is incorporated into the dialog,

<sup>8</sup><http://www.liwc.net>

i.e., the agent generates questions which explicitly refer to the displayed emotional facial expressions and their relevance for the affective states expressed verbally by a user.

For artificial conversational systems the inability to generate a suitable response on the basis of an analysis (i.e., semantic, discourse) of recent utterances is a relatively frequent problem, especially in open-domain applications. The third initiation condition for the application of affective cues in the dialog management provides the possibility to shift, in such situations, the focus of the system response from the semantics of the user utterance, to e.g., its affective content. In the current system the list of categories used for applying this method of response generation includes: positive or negative emotion, swear words, anger and health. They are detected using the LIWC dictionary.

#### A. Affect Bartender AIML set

The Affect Bartender AIML set (AB-AIML) provides a robust fall-back mechanism capable to generate system responses for a range of inputs which do not match activation cues of the provided ALDS scenarios, cf. section V-B. In particular it generates response candidates for open domain contexts. The adaptation of a more generic Affect Listener AIML<sup>9</sup> set [26] for the purposes of the Affect Bartender system was twofold, aiming at enabling the system to generate response candidates that:

- convey the Virtual Bartender’s openness, interest in users’ feelings, current mood, events which are of importance for them, etc.
- provide knowledge specific to the bartender tasks, and the virtual bar settings.

The response candidates are acquired based on the AIML response retrieval mechanism. The AB-AIML set contains 14825 patterns, 8549 response instructions, 782 ‘that’ statements and 6999 ‘srai’ substitution rules.

#### B. Affect Listeners Dialog Scripting

Affect Listeners Dialog Scripting (ALDS), described in more detail in [26], enables the creation of interaction scenarios that:

- provide capabilities to manage task oriented parts of verbal communication that include several dialog turns, i.e., system and user utterances,
- take advantage of the system’s perception capabilities (i.e., natural language analysis, affective states analysis) that extend beyond a simple matching mechanism that is solely based on keywords or textual pattern.

In particular, in the presented version of the Affect Bartender system, the ALDS scenarios rely on the affective, linguistic and cognitive categories discovered in a user utterance. In contrast to more complex communication tasks, e.g., receiving orders in a virtual bar context, the usage of affective cues relies on a pre-defined link between an initiation condition (e.g., user inputs and/or system state) and a specific

---

User: jeez.. that sucks!

Perception Layer: DA-Statement LC:Swear: NS- -4  
Dialog State (match. cond.): LC-Swear,  
Confusion\_statement=1

System: most of our clients don’t use this type of expressions. i would appreciate if you don’t use them too often also ;) thanks!

---

Fig. 3. Dialog excerpt based on AB-ALDS - LIWC activation cue applied in the Confusion Statement state.

system response template. For example, information about an extraordinarily high positive or negative sentiment in a user utterance can trigger a system response focusing on the expressed state rather than on the content of a user utterance e.g.: negative sentiment (NS= -5), example system response: “are you disappointed? ..if it is my fault, i am really sorry..”.

The ALDS scenario used in the Affect Bartender system also incorporates information on dialog act classes, surface features of the utterances (e.g., emoticons) and the base valence and arousal values used for the generation of emotional facial expressions for the avatars and VH. The last feature enables to incorporate coarse-grained information about the facial expression observed in a given moment by a user. For examples, if an utterance is classified as negative and at the same time the valence expressed by the avatar’s facial expression is positive the system might react by generating an utterance, such as: “You are smiling, but what you say doesn’t sound that optimistic... What is going on?”.

Further, rapid changes in the base valence values for user’s avatar facial expressions can trigger the system with a response that relates to that. For instance if a difference between two consecutive valence values assigned to the avatar’s emotional facial expressions is significant ( $n > 4$ ), and during both utterances the arousal value {a} was larger than a set threshold, the system might generate the following response candidate: “You look much more optimistic than just a second before... Gooooood :) Can you tell me what cheered you up?”.

Response generation scenarios that are based on the detected LIWC categories are applied to create system utterances in situations when all of the so-far generated response candidates contain a “confusion statement”. This signals the system’s inability to relate to the semantic content of a user utterance. For example, the discovery of swear words in a user utterance might lead in such cases to the following system response: “Most of our clients don’t use this type of expressions. I would appreciate if you don’t use them too often also ;) Thanks!”.

Figure 3 presents an example along with the activation cues discovered in a user utterance annotated in the Perception Layer.

Each category of the affect-related ALDS scenarios (i.e., sentiment class, emotional facial expressions, particular affective or linguistic category of LIWC dictionary) is used only once during an interaction.

<sup>9</sup>AIML - Artificial Intelligence Markup Language

---

System: so what would you like to have?  
 Dialog State (update): Accepting\_Orders=1  
 User: do you happen to have sidecar?

---

Perception Layer: DA-ynQuestion DR-sidecar  
 Dialog State (match. cond.): Accepting\_Orders=1,  
 drink-1, DA-Order=0

---

System: yes, we have it.. would you like it now?  
 User: yes, please.

---

Perception Layer: DA-Order  
 Dialog State (match. cond.): Accepting\_Orders=1,  
 drink-1, DA-Order=1

---

System: ok :-) sorry it took a while...  
 here you are.. enjoy!  
 User: thanx!

---

Fig. 4. Dialog excerpt based on AB-ALDS - example of order reception.

The management of tasks, specific for a bartender relies on the system’s capability to detect particular cues in user utterances (i.e., Perception Layer annotation) and relate these cues to the observed dialog states. Figure 4 presents an excerpt of a system-user communication, typical for the initial stages of dialogs conducted with the experiment participants. The provided ALDS rules enable the system to update the dialog states (e.g., system response: “So what would you like to have?” results in the activation of an “Accepting\_Orders” state) and to progress a dialog in an attempt to discover the remaining activation cues that enable a completion of a particular dialog scenario. As long as the particular dialog state is active, the system stores the scenario’s activation cues (e.g., detection of a drink name in a user utterance – “sidecar”; detection of a particular Dialog Act class – “Order”). If all the required activation cues, necessary for a completion of a scenario are discovered, the system generates a closing response (e.g., “ok :-) sorry it took a while... here you are.. enjoy!”). In such case, the system’s dialog memory is cleared from the activation cues that were relevant to the recently completed scenario and are not related to other active scenarios.

Figure 5 presents an excerpt of the Affect Bartender - user interaction log, recorded during experiments in a Virtual Reality setting that illustrates a change between a task specific part of a dialog (i.e., ordering drinks) with the one that aims at the detection of users affective states and the reasons for them.

## VI. CONCLUSIONS

Limitations of the input processing capabilities are often referred to as the restricting factor for a wider adaptation of virtual characters in various application scenarios and domains. This includes the ability to perceive affective and attentive cues from users and to appropriately incorporate this information to the system’s generated behaviour.

This work focuses on the presentation of methods for the

---

System: hi! what would you like to drink?  
 User: coke, please  
 System: here is your drink!... enjoy :D  
 User: thanks  
 System: you are quite welcome :) i feel i  
 should ask you how was your day so far?  
 User: it is not bad  
 System: is not bad?  
 User: no, i had a pretty good day today, why not  
 System: i see. any specific reasons for that?  
 User: actually, i got nice results for my experiments  
 ...

---

Fig. 5. Excerpt of the Affect Bartender - user communication at the beginning of an interaction.

perception of textual affective cues from system-user communication and their application for the management of conversational agent-user interactions. To date textual messages, either used alone or in combination with other modalities, are still the prevalent mode of communication between users and computer systems. In practical, real world applications, the dependency of computer systems on the textual modality provides an argument for the further research on methods for processing of the natural language input, including the detection and analysis of textual expressions of affective states. In numerous communication tasks, the ability of the computer systems to correctly identify the existence and polarity of emotions expressed by users, based on the analysis of short, informal messages is a prerequisite for the affective analysis of the ongoing communication and the basis for managing affective aspects of the interactions with users. Such tools also enable the annotation of textual communication between the users in open channels, e.g., blog discussions, Twitter-style short message exchanges, social networks status updates<sup>10</sup>, which provides insights on the role of emotions in the human-human communication. These are useful for modelling, and in future, for application of the acquired insights to the next generation of affective conversational agents.

## ACKNOWLEDGMENT

The work was supported by a European Union grant in the 7th Framework Programme, Theme 3: Science of complex systems for socially intelligent ICT. It is part of the Cyber-Emotions project (contract 231323). The work presented here builds on joint work within the CyberEmotions project. In particular the authors would like to thank our colleagues from the following institutions: EPFL, Jacobs University, Warsaw University of Technology, ETH Zurich, Jozef Stefan Institute, TU Berlin and Gemius SA.

<sup>10</sup><http://myspace.com>



## REFERENCES

- [1] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? sentiment classification using machine learning techniques," in *Proceedings of EMNLP 2002*, 2002.
- [2] J. Blitzer, M. Dredze, and F. Pereira, "Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification," in *Proceedings of ACL '07*, pp. 440–447.
- [3] M. Thomas, B. Pang, and L. Lee, "Get out the vote: determining support or opposition from congressional floor-debate transcripts," in *Proceedings of EMNLP '06*, 2006, pp. 327–335.
- [4] A. Devitt and K. Ahmad, "Sentiment polarity identification in financial news: A cohesion-based approach," in *Proceedings of ACL '07*, pp. 984–991.
- [5] G. Mishne, "Experiments with mood classification in blog posts," in *1st Workshop on Stylistic Analysis Of Text For Information Access*, 2005.
- [6] A. Neviarouskaya, H. Prendinger, and M. Ishizuka, "User study on AffectIM, an avatar-based Instant Messaging system employing rule-based affect sensing from text," *International Journal of Human-Computer Studies*, vol. 68, no. 7, pp. 432–450, Jul. 2010.
- [7] M. Thelwall, K. Buckley, and G. Paltoglou, "Sentiment in short strength detection informal text," *Journal of the American Society for Information Science and Technology*, 2010. [Online]. Available: <http://onlinelibrary.wiley.com/doi/10.1002/asi.21416/full>
- [8] R. Picard, *Affective Computing*. MIT Press, Cambridge, MA, 1997.
- [9] A. Paiva, R. Prada, and R. Picard, "Affective computing and intelligent interaction," in *Second Int. Conf., ACHI 2007*, 2007.
- [10] M. Schroeder and R. Cowie, "Developing a consistent view on emotion-oriented computing," in *Machine Learning for Multimodal Interaction*, S. Renals and S. Bengio, Eds. Springer, 2006, pp. 194–205.
- [11] S. Yacoub, S. Simske, X. Lin, and J. Burns, "Recognition of emotions in interactive voice response systems," in *Proceedings of the 8th European Conference on Speech Communication and Technology (EUROSPEECH 2003)*, 2003, pp. 729–732.
- [12] H. Tang, S. M. Chu, M. Hasegawa-Johnson, and T. S. Huang, "Emotion recognition from speech via boosted gaussian mixture models," in *Proceedings of the 2009 IEEE international conference on Multimedia and Expo*, ser. ICME'09. Piscataway, NJ, USA: IEEE Press, 2009, pp. 294–297. [Online]. Available: <http://portal.acm.org/citation.cfm?id=1698924.1698997>
- [13] M. Danieli and E. Zovato, "The affective dimension of speech acts and voice expressiveness," *Spoken Communication*, pp. 191–204, 2010.
- [14] N. Sebe, M. Lew, Y. Sun, I. Cohen, T. Gevers, and T. Huang, "Authentic facial expression analysis," *Image and Vision Computing*, vol. 25, no. 12, pp. 1856–1863, 2007.
- [15] R. Cowie, E. Douglas-Cowie, K. Karpouzis, G. Caridakis, M. Wallace, and S. Kollias, "Recognition of emotional states in natural human-computer interaction," in *Multimodal User Interfaces*, ser. Signals and Communication Technology, D. Tzovaras, Ed. Springer Berlin Heidelberg, 2008, pp. 119–153.
- [16] S. DMello, R. Picard, and A. Graesser, "Toward an affect-sensitive autotutor," *IEEE Intelligent Systems*, vol. 22, no. 4, pp. 53–61, 2007.
- [17] Z. Zeng, J. Tu, M. Liu, T. S. Huang, B. Pianfetti, D. Roth, and S. Levinson, "Audio-visual affect recognition," *IEEE Transactions on Multimedia*, vol. 9, p. 2007, 2007.
- [18] C. Pelachaud and I. Poggi, "Towards believable interactive embodied agents," in *Fifth Int. Conf. on Autonomous Agents workshop on Multimodal Communication and Context in Embodied Agents*, 2001.
- [19] P. Gebhard, M. Schroeder, M. Charfuelan, C. Endres, M. Kipp, S. Pammi, R. M., and T. O., "Ideas4games: Building expressive virtual characters for computer games," in *In: Proceedings of the 8th International Conference on Intelligent Virtual Agents*, ser. LNAI. Springer, 2008, pp. 426–440.
- [20] N. Reithinger, P. Gebhard, M. Loeckelt, A. Ndiaye, and M. Pflieger, N. and Klesen, "Virtualhuman - dialogic and affective interaction with virtual characters," in *In: Proceedings of the 8th International Conference on Multimodal Interfaces*, 2006.
- [21] G. Caridakis, A. Raouzaiou, E. Bevacqua, M. Mancini, K. Karpouzis, M. L., and C. Pelachaud, "Virtual agent multimodal mimicry of humans," *Language Resources and Evaluation, special issue on Multimodal Corpora For Modelling Human Multimodal Behavior*, 2008.
- [22] E. Andre, L. Dybkjaer, W. Minker, and P. E. Heisterkamp, "Affective dialogue systems, tutorial and research workshop (ads 2004)," in *Lecture Notes in Computer Science*. Springer, 2004.
- [23] J. Pittermann, A. Pittermann, and W. Minker, *Handling Emotions in Human-Computer Dialogues*. Dordrecht (The Netherlands): Springer, 2009. [Online]. Available: <http://www.springer.com/linguistics/computational+linguistics/book/978-90-481-3128-0>
- [24] K. Bhatt, S. Argamon, and M. Evens, "Hedged responses and expressions of affect in human/human and human/computer tutorial interactions," in *Proceedings of the 26th Annual Conference of the Cognitive Science Society (CogSci04)*, 2004, pp. 114–119.
- [25] B. Reeves and C. Nass, *The media equation: how people treat computers, television, and new media like real people and places*. Cambridge University Press, 1996.
- [26] M. Skowron, "Affect listeners. acquisition of affective states by means of conversational systems," in *Development of Multimodal Interfaces - Active Listening and Synchrony*, ser. Lecture Notes in Computer Science. Springer, 2010, pp. 169–181.
- [27] E. Forsyth and C. Martell, "Lexical and discourse analysis of online chat dialog," in *Proc. of the First IEEE Int. Conf. on Semantic Computing*, 2007, pp. 19–26.
- [28] E. T. Jaynes, "Information theory and statistical mechanics," *Physical Review*, vol. 106(4), pp. 620–630, 1957.
- [29] J. W. Pennebaker, M. E. Francis, and R. K. Booth, *Linguistic Inquiry and Word Count: LIWC 2001*. Erlbaum Publishers, 2001.
- [30] M. Skowron, J. Irran, and B. Krenn, "Computational framework for and the realization of cognitive agents providing intelligent assistance capabilities," in *European Conference on Artificial Intelligence. The 6th Int. Cognitive Robotics Workshop*, 2008, pp. 88–96.
- [31] S. Gobron, J. Ahn, S. Quentin, D. Thalmann, M. Skowron, S. Rank, G. Paltoglou, M. Thelwall, and A. Kappas, "3d-emochatting: an interdisciplinary communication model for vr chatting," in *review process*.
- [32] G. Paltoglou, S. Gobron, M. Skowron, M. Thelwall, and D. Thalmann, "Sentiment analysis of informal textual communication in cyberspace," in *In Proc. Engage 2010, Springer LNCS State-of-the-Art Survey*, pp. 13–25.
- [33] J. W. Pennebaker, M. R. Mehl, and K. Niederhoffer, "Psychological aspects of natural language use: our words, our selves," *Annual Review of Psychology*, vol. 54, pp. 547–577, 2003.
- [34] C. K. Chung and J. W. Pennebaker, "Revealing dimensions of thinking in open-ended self-descriptions: An automated meaning extraction method for natural language," *Journal of Research in Personality*, vol. 42, pp. 96–132, 2008.