An NVC Emotional Model for Conversational Virtual Humans in a 3D Chatting Environment

Junghyun Ahn[1], Stéphane Gobron[1,2], David Garcia[3], Quentin Silvestre[1], Daniel Thalmann[1,4], and Ronan Boulic[1]

[1] Immersive Interaction Group (IIG), EPFL, Lausanne, Switzerland
[2] Information and Communication Systems (ISIC), HE-Arc, St-Imier, Switzerland
[3] Chair of Systems Design (CSD), ETHZ, Zurich, Switzerland
[4] Institute for Media Innovation (IMI), NTU, Singapore

Abstract. This paper proposes a new emotional model for Virtual Humans (VHs) in a conversational environment. As a part of a multi-users emotional 3D-chatting system, this paper focus on how to formulate and visualize the flow of emotional state defined by the Valence-Arousal-Dominance (VAD) parameters. From this flow of emotion over time, we successfully visualized the change of VHs' emotional state through the proposed *emoFace*s and *emoMotion*s. The notion of Non-Verbal Communication (NVC) was exploited for driving plausible emotional expressions during conversation. With the help of a proposed interface, where a user can parameterize emotional state and flow, we succeeded to vary the emotional expressions and reactions of VHs in a 3D conversation scene.

Keywords: Emotional model and dynamics, facial expressions.

1 Introduction

Effective control and visualization of Non-Verbal Communication (**NVC**) is essential for development of a plausible industrial [1,2] or academic [3,4] application, involving 3D social communication. As illustrated in Figure 1, the NVC consists of complex emotional events such as facial expression, head orientation, gaze, emotional body motion, breathing, and etc. Since a large part of these emotional information is transmitted in an unconscious manner, it is not easy to be fully controlled by a user. Moreover, a Virtual Human (**VH**) acting as a user's avatar or conversational agent involves understanding of emotion, which is also not an easy task as this concept is shown to be difficult to define [5,6]. In this context, we propose an efficient emotional model that facilitates user's control for visualizing affective conversational VHs.

The basic emotional state of VH is interpreted by using three dimensional emotion axes, **V**alence, **A**rousal, and **D**ominance (**VAD**) [7,8,9]. These VAD parameters simplify task of designing user interface and easily transfers emotional states through conversational VHs. From our proposed emotional model and dynamics, the VAD parameters vary over time and visualize the NVC attributes from interactive emotional events. A <u>*demo video*</u> is provided for more detail of each process described in this paper (see also Section 5).

F.J. Perales, R.B. Fisher, and T.B. Moeslund (Eds.): AMDO 2012, LNCS 7378, pp. 47–57, 2012.

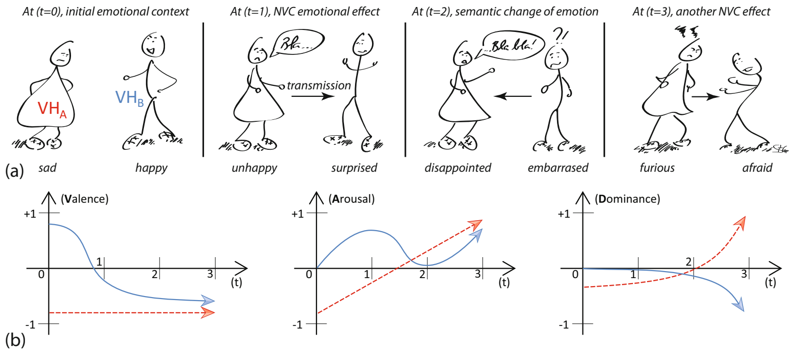

Fig. 1. (a) time-line of complex emotion variations highlighting NVC; (b) corresponding VAD variations

2 Related Works

A number of researches on conversational VH have been presented so far. A behavior expression animation toolkit (BEAT) that allows animators to input typed text to be spoken by an animated human figure was introduced [10]. Later, a dialog agent "Max" was developed as museum guide [11] and a gaming opponent [12]. Two virtual agents was also presented in an interactive poker game [13]. They applied the RFID technique to understand real world situation with speech recognition. Recently, two museum guide agents "Ada" and "Grace" [14] and a bartender agent was also introduced within an interdisciplinary VR architecture [15]. The VA and VAD parameters were also applied in [16,15] and [11,12]. However, they mainly focused on dialog agents, system design, speech recognition, and a single user environment. The transmission of avatar's emotion in a multi-users scene was not considered.

A model of behavior expressivity using a set of six parameters that act as modulation of behavior animation has been developed [17] as well. Recently, a constraint-based approach to the generation of multi-modal emotional displays was introduced [18]. Those works offer a good insight about how to derive sequential behavior including emotions. However, our approach gives more weight to the NVC emotional dynamics and its visualization in a conversational scene.

3 Emotional Model

The emotional face and body expressions influence the whole conversation even if a spoken sentence is short. We have developed a model where a single utterance produces continuous changes of emotion as depicted in Figure 1.

3.1 Types of Emotions

To parameterize the emotion of a **VH**, we choose the **VAD** coordinate system, as it allows any emotions to be simulated. The corresponding flow of emotion is influenced by users' typed sentences, users' direct emotional command, memory towards a VH, time attenuation functions, direction of dialog, and etc. Therefore, in the context of social communication, at least three aspects of emotions have to be taken into account: (1) short term emotions; (2) long term emotions; and (3) inter-personal emotions (*i.e.* individual opinion towards encountered VHs). These aspects are presented in the following paragraphs.

Short Term Emotions (STE). We process the messages (typed by a user, generated by a computer, or scripted) to extract information about the most recent emotional content in terms of polarity and VAD. Each VH taking part of the conversation is influenced by this extracted information. Moreover, users' direct VAD input also influences the change of emotional flow. As the core of the proposed model, more details on STE (*emotional dynamics*) will be explained later in Section 4. Facial expressions generated from VAD value (see Sub-section 3.2) and breathing rhythm from the arousal are the main outputs generated by STE.

Long Term Emotions (LTE). The model of internal emotion dynamics provides us with a data-driven estimation of the LTE of a user, given the set of STE expressed and perceived by the user. Emotional body motions (see Subsection 3.3) is triggered from the interpretation of this LTE.

Inter-Personal Emotions (IPE). Another type of emotion is related at individual level, *i.e.* how "we" appreciate people at individual scale. To simulate this aspect, we memorize all conversations received from every individual separately, and we keep updated a special variable that stores the current state of the emotions between each pair of VH. Then, if we want to know the emotional appreciation that a specific VH has towards another, we can access this variable, which relaxes on time the same way as internal emotions do. Accumulated average VAD toward a VH is added as an offset to the current STE VAD state.

We believe that this classification can be also useful to simulate *autonomous agents* with emotions.

3.2 Facial Expressions: *emoFace*s

Based on an internal survey (11 subjects, one woman and 10 men from 25 to 43 years old), we derived 27 emotional words potentially corresponding to the VAD space as presented in Figure 2. We defined three different areas, *i.e.* negative, neutral, and positive for each axis and attribute 27 emotional adjectives for each spatial region. Those words were referred to analyze facial expressions (*emoFace*s) and emotional body movements (*emoMotion*s).

As a main output of the STE, facial expression is continuously updated by the VAD value obtained from each simulation loop. Therefore, it was necessary to design a framework that generates *emoFace*s from any VAD coordinate point. In this context, we exploited 7 face parts and 11 face-part actions (see Table 1), which are extracted from Facial Action Coding System (FACS)'s Action Unit (AU)s [19] and Facial Animation Parameter (FAP) [20]. We cross referenced each face-part actions with the emotional words in the VAD model and exploited

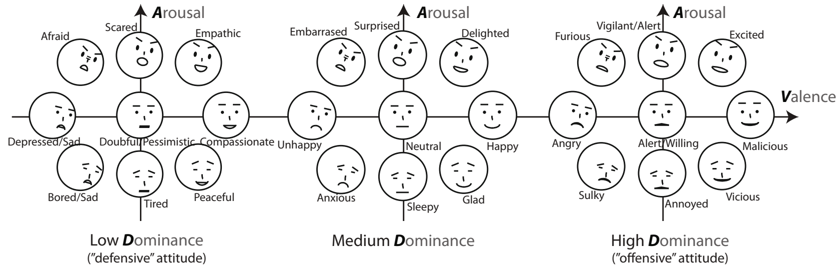

Fig. 2. A VAD model with emotional terms and its facial expressions

the dependency of each action on emotional axes as presented in Table 1. Each facial expression of emotional words has been explored from our internal observation and related researches in social psychology [21,22,23]. Besides expressions, the *emoFace* also controls a simple text-to-visual-speech and head turn motions.

Table 1. Dependency of each face-part action on **VAD** axes. Related axis to the action is mark with an 'o'. Detailed *emoFace*s are illustrated in Figure 2.

Face parts	Face-part actions	Related AUs (FAPs)	**V**	**A**	**D**
1. Eyebrows	Tilt left and right	AU 1,2 (FAP 31,32,35,36)		o	
	Move up and down	AU 4 (FAP 33,34)	o	o	
2. Eyelids	Move up and down	AU 5,7 (FAP 19,20)		o	
3. Upper lip	Move up and down	AU 10			o
4. Lip corners	Tilt left and right	AU 12,15 (FAP 12,13)	o		
	Stretch and Pull	AU 18,20 (FAP 6,7)	o		o
5. Mouth	Open and Close	AU 17,26		o	o
6. Head	Turn left or right	AU 51 or 52	o		o
	Head up and down	AU 53,54		o	
	Tilt left or right	AU 55 or 56	o	o	
7. Eye	Look at	AU 61-64 (FAP 23-26)		o	o

3.3 Emotional Body Movements: *emoMotion*s

From the emotional words defined in Figure 2 and with the participation of a professional actor, we associated emotional body movements to fit all 27 areas of the VAD model. Each full-body motion (including fingers movement and facial animation) was captured around 5 seconds.

Unlike *emoFace*, and as an output of LTE, an *emoMotion* is triggered from the analysis of emotions over a longer temporal window (*e.g.* 10 seconds). A recently peaked VAD area is interpreted as a recent emotion and is updated in each simulation loop. The peak is measured by the norm of VAD and the *emoMotion* activates when the current VAD hits again this area.

Another type of body movements, the breathing motion is related to STE rather than LTE. It shares the control of five joints (left/right sternoclavicular, left/right acromioclavicular, and head) during the whole simulation loop. Therefore, any activated motion such as *emoMotions*, conversational motion, or etc is combined with the background breathing motion, which it adds an angular contribution given by $\sin(\omega t + \phi)$, where t is the current simulation time. The angular frequency is $\omega = (a+1)^3+1$, where a is the current arousal value, resulting in a range of breathing frequencies within [0.15, 1.43] *Hz*. Then, the phase is updated by an equation $\phi = \phi + \omega \Delta t$, for the purpose of smoothing the sine curve upon change of ω. Figure 3 shows examples of *emoFaces* and *emoMotions*.

Fig. 3. The effect of changing one dimension applied to *emoFaces* and *emoMotions*. The *emoFaces* need exact VAD coordinate, since each VAD state has different facial expression. The *emoMotions* are presented with emotional words depicted in Figure 2.

4 Short Term Emotions (STE)

The STE model contains two different types of input from the users: text message (S_i) and an external VAD (U_i) input, which is also described in the companion paper [24]. The S_i is processed to extract three types of values: (1) sentiment polarity from a lexicon-based classification [25]; (2) VAD from averaged ANEW values [26]; and (3) target of the emotion in terms of grammatical person ("I", "you", "it") from LIWC word frequency analysis [27].

The *emoFaces* and *emoMotions* are driven by the internal state of the *agent* as part of a data-driven simulation of our agent-based model, which is designed within the framework for collective emotions [28,29]. Therefore, in this section, we use the *Systems Design* notation of the term *agent* [30] instead of the term VH. Each *agent* has an emotional state composed of the VAD variables: v_i, a_i, and d_i. Each one of these variables x_i of *agent* A_i changes continuously over time through an equation of the form $\dot{x_i} = -\gamma_x x_i + \mathcal{F}_x(x_i, h)$.

The first term models an exponential decay of parameter γ_x that makes the emotional states relax to the neutral in the absence of input. The second term models the effect of communication and user input on the variable, given the communication fields present in the simulation, as explained below. Each one of these fields is subject to a decay of the same type as the emotional states of the *agents*, of the form $\dot{h} = -\gamma_h h$. This decay of the field represents the aging of emotional information and the vanishing memory of the system.

The first process that defines the changes in the internal state of an *agent* is a refinement process based on the user input and messages. This type of information is stored in the self-influence field (*VAD field*) h_{ii} of *agent* A_i, with a component for each of the VAD axes. This individual field receives the discrete information of the user expression and influences its internal state in a smooth manner. When a user provides a S_i or U_i, the value of each one of this fields is updated to a weighted average between its previous value and the input. This averaging depends on the input type and follows the equation $h = (1-s_y)h+s_y Y_i$, where Y_i can be S_i or U_i. The weight of the change in the field s_y will depend on the input, being s_u for U_i, s_s for S_i classified as 2^{nd} or 3^{rd} person, or $s_i > s_s$ if the message is classified as 1^{st} person.

The second process is the simulation of changes in emotional states due to message exchanges in the system. This information, extracted from the polarity classification of the messages, aggregates in two different kinds of fields, related to conversation and social identity (*signed field*).

The conversation field $h_{ij\pm}$ between the *agents* A_i and A_j has two independent components for positive and negative information. Each component relaxes independently, and is affected by the messages (with an influence factor s_d) coming from the conversation *agents* through equation $\dot{h}_\pm = -\gamma_h h_\pm + s$.

The social identity field $h_{i\pm}$ keeps the importance of *agent* A_i in the community, aggregating all the information related to each conversation it has been involved in. The size of the change in these conversation fields will depend on whether the messages taken as input were classified as 2^{nd} person (s_t), being then larger than if they were classified as 1^{st} or 3^{rd}.

Valence of an *agent* is affected by the conversation fields with an asymmetric effect depending on its sign. The second change comes from the self-influence fields, forcing the attractors to particular values stated by the user or inferred from the expression. When $v \geq 0$, the $\mathcal{F}_v$ is defined as Equation 1. Otherwise, when $v < 0$, h_{ij+} and h_{ij-} are switched. The attention parameter α defines the asymmetry of positive and negative emotions, b_0 is the bias towards positive emotions and b_3 is the saturation that avoids infinite valences. β_v is the strength factor of the update to known values of the state from utterances or user input, compared with the rest of the dynamics. When $v = 0$, the term $sgn(v)$ is 1.

$$\mathcal{F}_v = (\alpha h_{ij+} + (1-\alpha)h_{ij-})(b_0 \ sgn(v) - b_3 v^3) + \beta_v(h_{iiv} - v) \qquad (1)$$

Arousal increases with all the information present in the system, regardless of its valence but depending on how relevant it is for the individual. This way, the arousal will increase with all the fields and decrease only based in relaxation or internal assessments of low arousal, coming from the S_i or U_i. The parameter η balances how stronger is the identity field rather than the conversation, and d_0 and d_1 work similarly as their valence counterparts but with a linear saturation. The weight β_a is the importance of the self-field as for the valence.

$$\mathcal{F}_a = ((1-\eta)h_{ij} + \eta h_i)(d_0 - d_1 a) + \beta_a(h_{iia} - a) \tag{2}$$

Dominance dynamics are based on the social identity fields, and how the information directed to the individual changes its power regarding emotions. The parameters g_+ and g_- represent the asymmetric effect on the dominance, if a fearful reaction is supposed to be fast, $g_- > g_+$.

$$\mathcal{F}_d = g_+ h_{i+} - g_- h_{i-} + \beta_d(h_{iid} - d) \tag{3}$$

5 Parameterizations and Results

The proposed emotional model includes various parameters and variables, which define the emotional state and enable to differentiate the flow of emotion. In this section, we describe how we parameterized these values and show the result.

Table 2. A description of the short term emotions parameters and their values

Category	Par.	Description	Val.	Category		Par.	Description	Val.
Valence ($\mathcal{F}_v$)	α	attention shift	0.5	**D**ominance ($\mathcal{F}_d$)		$g+$	pos influence speed	0.7
	b_0	direct influence	0.9			$g-$	neg influence speed	0.7
	b_3	saturation	0.9			β_d	update strength	0.8
	β_v	update strength	0.8			γ_d	decay	0.1
	γ_v	decay	0.1	Field		γ_h	general decay	0.1
Arousal ($\mathcal{F}_a$)	η	identity strength	0.5		Signed ($\dot{h}_\pm$)	s_d	discussion influence	0.4
	d_0	direct influence	0.8			s_t	targeted influence	0.4
	d_1	saturation	0.8		VAD (h)	s_s	baseline influence	0.3
	β_a	update strength	0.8			s_i	"I" influence	0.5
	γ_a	decay	0.1			s_u	user input influence	0.8

Twenty different parameters (see Table 2) are assigned for a VH, and through the parametrization, we could vary the emotional attributes of each conversational VH. The influence functions $\mathcal{F}$ of each VAD can be parameterized independently and different social behaviors can be observed from the parametrization of signed and VAD fields. For example, the increment of decays causes a VH to lose its emotional feelings more or less quickly. The decrement of update strengths makes a VH

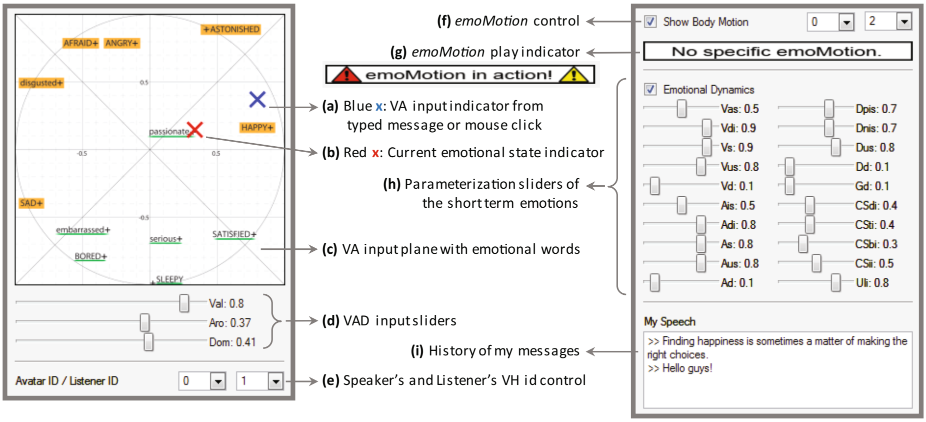

Fig. 4. The UI for emotional VH control. A message "Finding happiness is sometimes a matter of making the right choices" was typed in from a user. (a) and (d) show the goal VAD state {0.80,0.37,0.41} extracted from the typed message. The detail of (h) short term emotions parameters are described in Table 2.

Fig. 5. Three-VH conversation simulated and rendered by our proposed model

less emotional from the coming events. According to our observation, parameter values in Table 2 simulated a plausible animation of VHs conversation. These STE parameters was controlled by a User Interface (UI) introduced in Figure 4.

Our simulated conversation scene is illustrated in Figure 5 and the details of our results can be seen at the following link: http://youtu.be/UGbW8nDNO24. Another demo video of our companion paper can also be found *here*.

6 Conclusion

In this paper, we presented a new emotional model for text-based conversational environment. This approach adds a new dimension to animation applicable to virtual worlds where resulting conversations enable to create more natural direct and indirect exchanges. Original contributions introduced in this paper were: (a) multiple levels of emotion for each VH (STE, LTE, and IPE); (b) the *emoFace*s and *emoMotion*s associated with VAD model.

However, there are still some issues to be considered: First, the model can be improved by considering heterogeneous effects of *when* text utterances were exchanged. Second, the STE parameters have a potential to find interesting results on the personality of emotional VH. Third, in the case of avatar, this automatic method can be contested as it implies that the user and his/her avatar can have different opinions towards the same encountered VH. Nevertheless, when dealing with NVC, the user cannot control all possible parameters. The proposed direct VAD input could be served as a solution of this issue.

Acknowledgments. The authors wish to thank O. Renault, M. Clavien, and J. Conus for their collaboration on the mocap, modeling, and algorithm. This work was supported by a EU FP7 CYBER**EMOTIONS** Project (Contract 231323).

References

1. GUILE3D: Virtual assistant denise, `http://guile3d.com/en/products/`
2. Boellstorff, T.: Coming of Age in Second Life: An Anthropologist Explores the Virtually Human. Princeton University Press (2008)
3. 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 and spoken intonation for multiple conversational agents. In: SIGGRAPH 1994, pp. 413–420. ACM (1994)
4. Becker, C., Kopp, S., Wachsmuth, I.: Simulating the Emotion Dynamics of a Multimodal Conversational Agent. In: André, E., Dybkjær, L., Minker, W., Heisterkamp, P. (eds.) ADS 2004. LNCS (LNAI), vol. 3068, pp. 154–165. Springer, Heidelberg (2004)
5. Ekman, P.: Emotions revealed. Henry Holt and Company, LLC, New York (2004)
6. 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), 38–41 (2010)
7. Bush, L.E.: Individual differences multidimensional scaling of adjectives denoting feelings. Journal of Personality and Social Psychology 25(1), 50–57 (1973)
8. Russell, J.A.: A circumplex model of affect. Journal of Personality and Social Psychology 39(6), 1161–1178 (1980)
9. Scherer, K.R.: What are emotions? and how can they be measured? Social Science Information 44(4), 695–729 (2005)
10. Cassell, J., Vilhjálmsson, H.H., Bickmore, T.: Beat: Behavior expression animation toolkit. In: SIGGRAPH 2001, pp. 477–486 (2001)

11. Kopp, S., Gesellensetter, L., Krämer, N.C., Wachsmuth, I.: A Conversational Agent as Museum Guide – Design and Evaluation of a Real-World Application. In: Panayiotopoulos, T., Gratch, J., Aylett, R.S., Ballin, D., Olivier, P., Rist, T. (eds.) IVA 2005. LNCS (LNAI), vol. 3661, pp. 329–343. Springer, Heidelberg (2005)
12. Becker, C., Nakasone, A., Prendinger, H., Ishizuka, M., Wachsmuth, I.: Physiologically interactive gaming with the 3d agent max. In: Intl. Workshop on Conversational Informatics, pp. 37–42 (2005)
13. Gebhard, P., Schröder, M., Charfuelan, M., Endres, C., Kipp, M., Pammi, S., Rumpler, M., Türk, O.: IDEAS4Games: Building Expressive Virtual Characters for Computer Games. In: Prendinger, H., Lester, J.C., Ishizuka, M. (eds.) IVA 2008. LNCS (LNAI), vol. 5208, pp. 426–440. Springer, Heidelberg (2008)
14. Swartout, W., Traum, D., Artstein, R., Noren, D., Debevec, P., Bronnenkant, K., Williams, J., Leuski, A., Narayanan, 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: Safonova, A. (ed.) IVA 2010. LNCS, vol. 6356, pp. 286–300. Springer, Heidelberg (2010)
15. Gobron, S., Ahn, J., Silvestre, Q., Thalmann, D., Rank, S., Skoron, M., Paltoglou, G., Thelwall, M.: An interdisciplinary vr-architecture for 3d chatting with non-verbal communication. In: EG VE 2011: Proceedings of the Joint Virtual Reality Conference of EuroVR. ACM (September 2011)
16. Ahn, J., Gobron, S., Silvestre, Q., Thalmann, D.: Asymmetrical facial expressions based on an advanced interpretation of two-dimensional russell's emotional model. In: ENGAGE 2010 (2010)
17. Pelachaud, C.: Studies on gesture expressivity for a virtual agent. Speech Commun. 51(7), 630–639 (2009)
18. Niewiadomski, R., Hyniewska, S.J., Pelachaud, C.: Constraint-based model for synthesis of multimodal sequential expressions of emotions. IEEE Transactions on Affective Computing 2(3), 134–146 (2011)
19. Ekman, P., Friesen, W.V.: Facial action coding system. Consulting Psychologists Press, CA (1978)
20. Magnenat-Thalmann, N., Thalmann, D.: Handbook of Virtual Humans. Wiley (2004)
21. Ekman, P.: Universals and cultural differences in facial expressions of emotions. In: Nebraska Symposium on Motivation, pp. 207–283. University of Nebraska Press (1971)
22. Ekman, P.: About brows: Emotional and conversational signals. In: von Cranach, M., Foppa, K., Lepenies, W., Ploog, D. (eds.) Human Ethology, pp. 169–248. Cambridge University Press, Cambridge (1979)
23. Ekman, P., Friesen, W.V., Ancoli, S.: Facial signs of emotional experience. Journal of Personality and Social Psychology 39(6), 1125–1134 (1980)
24. Gobron, S., Ahn, J., Garcia, D., Silvestre, Q., Thalmann, D., Boulic, R.: An Event-Based Architecture to Manage Virtual Human Non-Verbal Communication in 3D Chatting Environment. In: Perales, F.J., Fisher, R.B., Moeslund, T.B. (eds.) AMDO 2012. LNCS, vol. 7378, pp. 58–68. Springer, Heidelberg (2012)
25. Paltoglou, G., Gobron, S., Skowron, M., Thelwall, M., Thalmann, D.: Sentiment analysis of informal textual communication in cyberspace. In: ENGAGE 2010, pp. 13–15 (2010)
26. Bradley, M.M., Lang, P.J.: Affective norms for english words (ANEW): Stimuli, instruction manual and affective ratings. Technical report, The Center for Research in Psychophysiology. University of Florida (1999)

27. Pennebaker, J., Francis, M., Booth, R.: Linguistic Inquiry and Word Count, 2nd edn. Erlbaum Publishers (2001)
28. Schweitzer, F., Garcia, D.: An agent-based model of collective emotions in online communities. The European Physical Journal B - Condensed Matter and Complex Systems 77, 533–545 (2010)
29. Garcia, D., Schweitzer, F.: Emotions in product reviews empirics and models. In: IEEE SocialCom 2011 and PASSAT 2011, 483–488 (2011)
30. Schweitzer, F.: Brownian Agents and Active Particles. Collective Dynamics in the Natural and Social Sciences, 1st edn. Springer Series in Synergetics. Springer (2003)