Vis Comput (2010) 26: 505–519
DOI 10.1007/s00371-010-0446-x

ORIGINAL ARTICLE

From sentence to emotion: a real-time three-dimensional graphics metaphor of emotions extracted from text

Stephane Gobron · Junghyun Ahn · Georgios Paltoglou · Michael Thelwall · Daniel Thalmann

Published online: 7 April 2010

Abstract This paper presents a novel concept: a graphical representation of human emotion extracted from text sentences. The major contributions of this paper are the following. First, we present a pipeline that extracts, processes, and renders emotion of 3D virtual human (VH). The extraction of emotion is based on data mining statistic of large cyberspace databases. Second, we propose methods to optimize this computational pipeline so that real-time virtual reality rendering can be achieved on common PCs. Third, we use the Poisson distribution to transfer database extracted lexical and language parameters into coherent intensities of valence and arousal—parameters of Russell's circumplex model of emotion. The last contribution is a practical color interpretation of emotion that influences the emotional aspect of rendered VHs. To test our method's efficiency, computational statistics related to classical or untypical cases of emotion are provided. In order to evaluate our approach, we applied our method to diverse areas such as cyberspace forums, comics, and theater dialogs.

Keywords Virtual reality · Distribution functions · Data mining · Text analysis · Psychology and sociology · Facial animation

1 Introduction

Visualizing a digital human's emotion is a well researched subject. Many researchers working in this field have achieved outstanding results in different research areas such as body motion, facial expression, cognitive modeling and gaze analysis. However, no previous study has dealt with the issue of a sentence-to-graphics real-time emotional representation in a context of a conversational environment.

Deriving emotional value from arbitrary textual input is an open and important problem to tackle as it can solve the interesting problem of how to automatically add emotional expression to virtual humans (avatars or agents). Many applications relative to virtual reality (VR) and entertainment can strongly benefit by giving textual emotions (see Fig. 1), e.g., human conversation (i.e., avatar vs. avatar), low-cost AI interaction (i.e., avatar vs. agent), or an embodiment of stored manuscripts. In order to achieve an emotional simulation, not only is a realistic animation technique necessary, but also a sophisticated sentiment analysis of the textual input.

Therefore, in this paper, we present a novel 3D visualization framework that extracts, refines, and renders a metaphor of emotion using text-based social interactive conversations in real-time. Our proposed approach can be classified into three main stages:

Emotion analyzer from textual input—Initially, the pipeline passes a given textual input through an emotion detection module. The module is comprised of two different classifiers: a lexicon-based and a language model classifier, whose aims are to provide an analysis of each sentence, in terms of the emotion that it conveys. The output is expressed in four distinct dimensions. The lexicon classifier return two numbers, one on a positive scale (from 1 to 5) and the other on a negative scale (from -5 to -1), which express the emotional strength of the words contained in the provided text. The language model classifier provides probabilities of how subjective/objective and positive/negative the text is.

S. Gobron (✉) · J. Ahn · G. Paltoglou · M. Thelwall · D. Thalmann
EPFL, IC ISIM VRLAB, Station 14, 1015 Lausanne, Switzerland
e-mail: stephane.gobron@epfl.ch

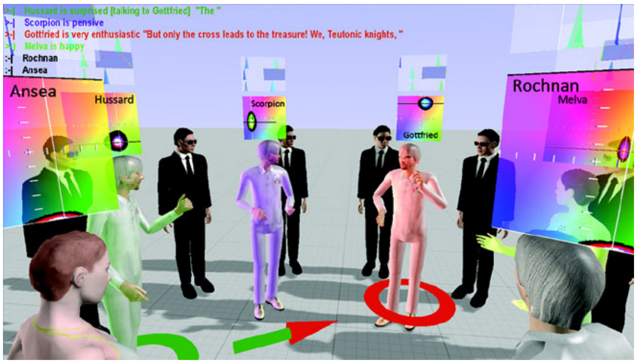


Fig. 1 Dialog scene with emotional statistic interpretation and the corresponding emotion graphical metaphor rendered on VH

Emotion generator by a probabilistic method—Using the output of the lexicon and language classifiers, we derive valence and arousal values to match the circumplex model [38]. Studies have indicated that the emotions and reactions of people during a conversation are nondeterministic (i.e., the same input can produce different levels of arousal and reaction). Therefore, a probabilistic approach is required to determine the current emotional state of each participant.

FACS-based emotion visualizer—The proposed probabilistic emotion generator gives one valence and one arousal value (ranging from -1.0 to 1.0). These parameters help generating a 2D colored map that guides the coordination of the emotional state of the VH. Each area of the VH face is controlled based on a so-called *FACS action unit* [13]. The valence and arousal values are used as weighting factors of the facial deformation and allow expressing an infinite number of expressions.

For every sentence of a textual conversation, our computational model and animation engine can process the emotion of the speaker in real time, i.e., extracting, refining, and rendering emotion parameters are almost transparent compared to the 3D rendering of the virtual human (which itself is greater than 60 fps).

The authors would like to stress that the notion of *emotions* is very relative to the way that human beings perceive them from one to another. The model presented in this paper aims to give the initial steps towards a statistically valid interpretation of instantaneous emotion. Therefore, historical background, personal characteristics, and social contexts are still not taken into account. Hence, when first testing our model, we made a mistake by thinking about a specific and personal feeling, then trying to find a corresponding sentence. With such a wrong approach, we were often unpleasantly surprised by the results from our model. Indeed, our model cannot know personal experience. Then, as we analyzed the result from a neutral sentence (i.e., being careful not to use personal experience), we observed that our model was surprisingly objective.

2 Background

2.1 Previous work relative to computer graphics

Research topics on computer graphics emotion can be classified into several research areas such as body motion, facial expression, cognitive modeling, gaze analysis, and text-to-graphics systems.

Body motion—Emotion influences the motion of the human body. Graphics researchers attempt to modify the captured or physically simulated motion according to emotion. Unuma et al. [43] proposed a method for modeling human figure locomotion with emotions by applying Fourier Expansions. Amaya et al. [2] calculated emotional transforms of the body motion from motion captured data. Costa et al. [11] applied Laban Movement Analysis (LMA) for the upper body motion by effort and shape components. Egges et al. [14] described idle motion generation by using PCA analysis. Stone et al. [39] proposed an integrated framework for creating interactive, embodied talking characters. Pelachaud [30] developed a model of behavior expressivity using a set of six parameters that act as a modulation of behavior animation. Different ways to control autonomous agents [10, 28, 41, 47] have been presented for the purpose of manipulating crowd movements. In this paper, we did not consider these body movements, however, applying them in our proposed framework should drastically improve our result.

Facial expression—There are numerous researches on facial expression related to emotion from different research fields such as VR, computer vision, computer animation, psychology, robotics, and human–computer interaction. Velasquez [44] shows different facial emotions among their experiments. Hess et al. [19] showed that happiness and fear expressions bias sex discrimination toward the female, whereas anger expressions bias sex perception toward the male. Pelachaud [29] defined a representation scheme that describes the temporal evolution of the expression of an emotion. Ekman and Friesen [13] proposed a Facial Action Coding System (FACS) with more than 40 action units (AU). Each action unit is related to the muscle movement and each unit is related to the facial emotion. Moreover, FACS has been proven to be useful to express facial emotion by psychologists and animators. Therefore, in this paper, we derived facial expression based on the FACS AU.

Cognitive modeling—Human emotion is a very complex matter that is derived from various external parameters [32]. For this reason, the *cognitive modeling* issue for human emotion has been researched and proposed. Becheiraz and Thalmann [5] proposed a social interaction and communication model between virtual agents. Cassell [6] and Badler et al. [4] proposed a way to design personality and emotion models of VH. Allbeck and Badler [1] described a parameterized action representation (PAR) that allows an agent to act or plan its own or other embodied agents' actions. Kshirsagar and Magnenat-Thalmann [23] also defined a multi-layered personality model based on the OCEAN model. Ruttkay et al. [37] exploited emotional change by defining a 2D emotion disc. Su et al. [42] predicted specific personality and emotional states from hierarchical fuzzy rules. For the previous works, emotional state transition, personality modeling, or research on social interaction were considered as ways to define the human brain for deriving emotion. Nevertheless, the brain itself is such a complex organ that an accurate prediction is almost impossible. Hence, in this paper we focus on the probabilistic approach, in order to simulate complex brain actions.

Gaze analysis—In the conversational environment, head movements help to express emotion in a realistic way [8]. Masuko and Hoshino [25] generated eye movements. Grillon and Thalmann [18] generated head movements.

Text-to-graphics system—Linguistic interpretation can greatly improve the automation of emotion-based systems. Perlin and Goldberg [31] proposed a system for creating a real-time behavior-based animated actor consisting of an animation and behavior engine. Cassell et al. [7] proposed the behavior expression animation toolkit (BEAT). This system allowed animators to input typed text to be spoken by an animated human figure. Coyne and Sproat [9] described linguistic analysis and depiction techniques needed to generate language-based 3D scenes. This allowed animators to input typed text to be spoken by an animated human figure. Our proposed framework mainly focused on the generation of infinite emotion expression from the textual input.

2.2 Opinion mining and sentiment analysis

The extraction of opinions and emotions from textual input, with an emphasis on blogs and fora, has enjoyed considerable popularity lately, both in academia and in industry. Two fields of research have addressed the above issue: opinion mining and sentiment analysis [34]. These aim at detecting whether segments of text (i.e., whole documents or sentences) contain opinions or report factual information. They also aim to detect whether an expressed opinion is positive or negative.

Addressing the problem of subjectivity detection, Hatzivassiloglou et al. [20] demonstrated that semantically oriented and gradable adjectives provide strong indicators of subjectivity. Wiebe [45] improved the approach by taking into consideration the distributional similarity of words, i.e., having extracted the linguistic relations between terms, their similarity was defined as the mutual information between the relation-word pairs. Pang et al. [34] approached the problem of sentiment analysis as a classification problem, using individual words as features and experimented with three different supervised classification algorithms (Support Vector Machines, Naive Bayes and Maximum Entropy [40]) and achieved the best results with the former, using binary features. More recently, Wilson et al. [46] used a lexicon of 8,000 subjectivity clues to first determine whether a phrase is subjective and then, using a classifier and different combinations of features, determined its polarity. Zhe and Boucouvalas [48] proposed an engine capable of analyzing input text from a chat environment, and extracting a simple model of emotion. Alm et al. [3] presented an empirically based prediction of text-to-emotion prediction. Our proposed approach directly handles the problem of text-to-emotion.

3 The overall architecture

Figure 2 summarizes the data framework presented in this paper to determine which facial expression and associated emotional color should be appropriate to generate a graphical metaphor for emotion. In this figure, the main real-time pipeline is shown with the three orange modules, where data are shown in blue, and the processing model in yellow. Notice that personal characteristic, social context, and historical background are not considered, but are discussed in Sect. 8. Here are the three main steps of our pipeline:

Extraction of emotion using data mining—From text communication to analyzing four lexical and language parameters (see Sect. 4).

Statistical refining of emotion—From negative/positive lexical interpretations N, P to an arousal intensity a, which is moderated by a statistical model. From both lexical N, P and language happiness h factor to a valence intensity v.

Graphics metaphor of emotion—From arousal parameter a and valence parameter v to facial animation that is associated with our proposed emotional color model.

4 Extracting emotions from text sentences

We view the problem of extracting the emotions from text sentences as a classification problem. The general aim of classification is, given a document[1] D and a fixed set of

[1] The term "document" is used here in the broadest of senses, signifying a sequence of words. In realistic environments, "document" can be any sort of textual communication between two or more parties, such as blog posts, forum comments or Instant Messaging utterances.

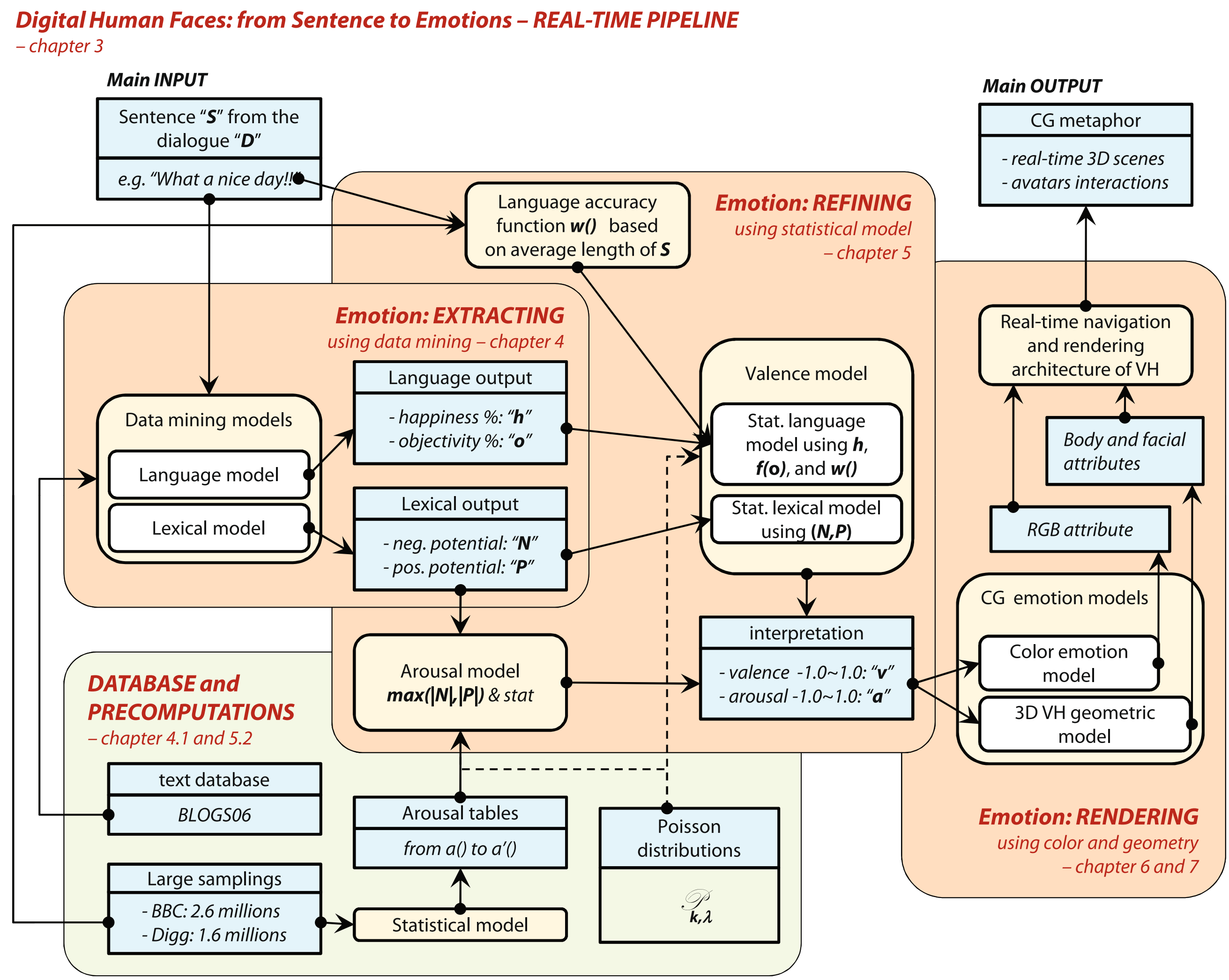


Fig. 2 Pipeline enabling real-time graphical metaphor of text emotions

classes $C = \{c_1, c_2, \ldots, c_t\}$, to assign D to one or more of the available classes.

Specifically, we approached the problem from two different perspectives that resulted in two different algorithms. Both are aimed at different classification problems and function independently of one another. The first one is an *unsupervised* lexicon-based classifier which utilizes various emotionally-annotated lexical corpora to extract the emotional polarity and intensity of the textual input, while the latter is a *supervised* Language model classifier.

4.1 Lexicon-based classifier

The lexicon-based classifier aims to extract the emotional content of a document D. Our approach is somewhat pertinent to the *rating-inference problem*, where the aim is to predict the author's evaluation of an object with respect to a multi-point scale (for example, film reviewers assigning 1 to 5 stars). Nonetheless, our approach significantly differs from that work in that it does not require training and does not produce a single outcome, but instead provides two independent ratings, one for the positive scale, $C_{\text{pos}} = \{+1, +2, +3, +4, +5\}$ and another for the negative scale, $C_{\text{neg}} = \{-1, -2, -3, -4, -5\}$, where bigger absolute values indicate stronger emotion and values $\{1, -1\}$ indicate its absence.

The algorithm is based on two dictionaries in order to extract the polarity and valence of terms. The first one is the General Inquirer (GI) lexicon from which we extracted the positive and negative word lists. The GI lexicon has often been used in research as the "gold standard" for algorithms that aim to automatically extract the sentimental orientation of words [34]. The second one is a dictionary based on the "Linguistic Inquiry and Word Count" (LIWC) software[2] which was derived from a number of psychological studies

[2] http://www.liwc.net.

and maintains an extensive dictionary along with human assigned emotional categories and strengths for each lemma.

Given a document D, the algorithm detects all words that belong to either dictionary and extracts their polarity and intensity. We enhance the initial term scores with additional functionalities such as negation detection (i.e., "good" versus "not good"), intensifiers (i.e., "liked" versus "liked very much") and diminishers ("excellent" versus "rather excellent") to produce the final document scores. The score of a document on the C_{pos} and C_{neg} scales is the maximum positive and negative number produced, respectively. We define those scores as P ("Positive") and N ("Negative").

4.2 Language model classifier

Additionally, we have also implemented two supervised classifiers for estimating the probabilities of whether a document D is objective or subjective, positive or negative. The classifiers function in a two-tier fashion. The first determines the probabilities of whether D is objective or subjective (i.e., $C_1 = \{obj, sub\}$), and the second-stage classification determines the probabilities of the polarity of the document (i.e., $C_2 = \{neg, pos\}$). We have utilized language model classifiers [27] for both classification tasks. The aim of the classifiers is to maximize the posterior probability $P(c|D)$, that a given document D belongs to class c. Typically, the best class is the *maximum a posteriori* (MAP) class c_{MAP}:

$$c_{\text{MAP}} = \arg\max_{c\in C}\{P(c|D)\}. \tag{1}$$

Using Bayes rule, we get

$$\begin{aligned} c_{\text{MAP}} &= \arg\max_{c\in C}\left\{\frac{P(D|c) * P(c)}{P(D)}\right\} \\ &\propto \arg\max_{c\in C}\{P(D|c) * P(c)\}. \end{aligned} \tag{2}$$

We have removed the denominator $P(D)$ since it does not influence the outcome of the classification. $P(c)$ is the prior that indicates the relative frequency of class c, i.e., all other things being equal, the classifier will prefer the most frequent class.

Language models operate by estimating the probability of observing document D, given class c. We represent D as token sequence $\{w_1, w_2, \dots, w_n\}$; therefore, the aim of a language model is to estimate the probability of observing the above sequence given c:

$$\begin{aligned} P(D|c) &= P(w_1, w_2, \dots, w_n|c) \\ &= P(w_1|c) * P(w_2|c, w_1) \\ &\quad * \cdots * P(w_n|c, w_1, w_2, \dots, w_{n-1}) \\ &= \prod_{i=1}^{n} P(w_i|c, w_1, \dots, w_{i-1}). \end{aligned} \tag{3}$$

Usually, an *n-gram* approximation is used to estimate (3), which assumes that the probability of token w_i appearing in document D depends only on the preceding $n-1$ tokens:

$$P(w_i|c, w_1, \dots, w_{i-1}) = P(w_i|c, w_{i-(n-1)}, \dots, w_{i-1}). \tag{4}$$

A straightforward way to calculate the maximum likelihood estimate of $P(w_i|c, w_{i-(n-1)}, \dots, w_{i-1})$ during the training phase of the classifier, given a set of documents and their respective categories, is by counting the frequency of occurrences of tokens sequences:

$$P(w_i|c, w_{i-(n-1)}, \dots, w_{i-1}) = \frac{\#(c, w_{i-(n-1)}, \dots, w_i)}{\#(c, w_{i-(n-1)}, \dots, w_{i-1})}, \tag{5}$$

where $\#(c, w_{i-(n-1)}, \dots, w_{i-1})$ is the number of occurrences of token sequence $w_{i-(n-1)}, \dots, w_{i-1}$ in documents of class c during the training phase and $\#(c, w_{i-(n-1)}, \dots, w_i)$ is respectively the number of occurrences of sequence $w_{i-(n-1)}, \dots, w_i$ in documents of class c during the training phase.

Despite the simplification that we introduced for the estimation of $P(D|c)$ using *n-grams* the probabilities of actually observing a significant number of times, any specific sequence of n tokens in the training set is susceptible to the *sparse data* problem, especially for large values of n (i.e., usually more than 3) and small to medium-sized training corpora. Simply put, high order n-grams do not occur often enough to provide a strong indication of class preference. For this reason, we usually further break the estimation of probability $P(D|c)$ to smaller n-grams, in a process that is called *smoothing*. There are various methodologies for smoothing, such as Laplace, Good–Turing or back-off estimators, but in this work we adopted the Witten–Bell approach. Therefore,

$$\begin{aligned} &P(w_i|c, w_{i-(n-1)}, \dots, w_{i-1}) \\ &\quad = \lambda_{(c,w_{i-(n-1)},\dots,w_{i-1})} * P(w_i|c, w_{i-(n-1)}, \dots, w_{i-1}) \\ &\qquad + (1 - \lambda_{(c,w_{i-(n-1)},\dots,w_{i-1})}) \\ &\qquad * P(w_i|c, w_{i-(n-2)}, \dots, w_{i-1}), \end{aligned} \tag{6}$$

where

$$\begin{aligned} &\lambda_{(c,w_{i-(n-1)},\dots,w_{i-1})} \\ &\quad = \frac{\#(c, w_{i-(n-1)}, \dots, w_{i-1})}{\#(c, w_{i-(n-1)}, \dots, w_{i-1}) + L * W_{(c,w_{i-(n-1)},\dots,w_{i-1})}}, \end{aligned} \tag{7}$$

with $W_{(c,w_{i-(n-1)},\dots,w_{i-1})}$ being the number of extensions of the specific token sequence, that is,

$$W_{(c,w_{i-(n-1)},\dots,w_{i-1})} = |\{w_k|\#(c, w_{i-(n-1)}, \dots, w_{i-1}, w_k) > 0\}| \quad (8)$$

and L being the *hyperparameter* of the distribution. Its aim is to provide a balance between higher and lower order n-grams. A high value of L gives more weight to lower n-grams, which is usually appropriate for smaller training sets, where the probability of encountering a higher order n-gram is small. In the reported experiments, the value of L was set to the value of the longest n-gram.

We experimented with unigrams and bigrams ($n = 1, 2$, respectively) which provide an acceptable compromise between effectiveness and efficiency. We trained the language model classifiers on the BLOGS06 dataset [26]. 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 three consecutive years by the *T*ext *RE*trieval *C*onferences (TREC).[3]

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: "1" if the document contains relevant, factual information about the entity but no expression of opinion, "2" if the document contains an explicit negative opinion towards the entity, and "4" if the document contains an explicit positive opinion towards the entity. We used the assessments produced from all 3 years of the conference, resulting in 200 different entity searches and 11,137 documents with a score of "1", 8,340 documents with a score of "2" and 10,457 with a score of "4", which were used as the "gold standard" for training our classifiers.

Specifically, for the first-level classification (i.e., $C_1 = \{obj, sub\}$) we used the documents that were given a value of "1" as objective and the union of "2" and "4" as subjective. For the second stage classifier (i.e., $C_2 = \{pos, neg\}$), we used the documents assigned a "2" as positive and "4" as negative. The resulting classifiers have an accuracy of approximately 70% for either classification task, using 10-fold cross-validation. We define the estimated probabilities $P(obj|D)$ as o ("objectivity") and $P(pos|D)$ as h ("happiness") for ease of reference from both classification tasks and pass them to the pipeline. Because our training set cannot be perfectly balanced, i.e., there is an unequal number of documents from each class, the initial probability estimates tend to have a strong preference for the most popular classes (see (1)). In order to remedy this behavior of the classifier and produce balanced probabilities, we used polynomial interpolation to produce the final probability estimates. Therefore, using t as the balance threshold: $h' = ah^2 + (1 - a)h$, with $a = (1 - 2t)/(2t^2 - 2t)$.

[3] http://www.trec.nist.gov.

5 Valence and arousal models

We first explain the Poisson distribution for computing the valence v and the arousal a emotion parameters, and second the models which are applied for both v and a.

Four parameters with different properties (i.e., range, intensities, means) are extracted from the text, all of them influencing in different ways the emotion we would like to deduce. Human emotions are particularly difficult to interpret, even for humans, and even more difficult to analyze in a VR context, for instance, fear of fire studied in a CAVE [24]. There is no universal well-defined model of emotion, and this paper does not pretend to solve the complexity of emotion or to make the perfect facial metaphor. We propose a bridge between text and graphics using a vector of the potential emotion. Concerning the relationship of emotions and facial expressions, most research has been made with discrete emotions [21, 36]. However, different concepts of emotions—discrete emotions vs. dimensional vs. appraisal—are linked to each other [12, 15].

The four values extracted from the text using classifiers represent possibilities or potentials rather than fixed sharp values with no error. Nevertheless, we still can derive hypotheses, and to do so, we must make a global model based on statistics. Furthermore, the use of discrete functions comes naturally as first, it allows pre-computations for fast rendering, and second, two of four parameters are already discrete numbers. For this purpose, the Poisson distribution [35] is one of the best mathematical tools: it returns probabilities of occurrences in a discrete domain, and multiple distributions can easily be combined.

Here is a short introduction to the equations used in our model. The Poisson distribution defines the discrete probability $\wp$ for a distribution of parameter λ so that the chance that k occurs with:

$$\wp_{k,\lambda} = \frac{\lambda^k e^{-\lambda}}{k!} = e^{\ln(\frac{\lambda^k e^{-\lambda}}{k!})} = e^{k \ln \lambda - (\lambda + \sum_{i=2}^{k} \ln i)}. \quad (9)$$

As factors λ^k and $k!$ can quickly reach the computational limit, we use the property that for any positive function $f = e^{\ln f}$.

The integral of the probabilities is always 1, but as λ increases, the maximum intensity of the resulting curve decreases and the wave shape becomes larger (see Fig. 3). This property is used to differentiate between the statistical occurrence of $|P|$ and $|N|$. As these absolute values are high (e.g., 5), the lexicon-based classifier indicates that the

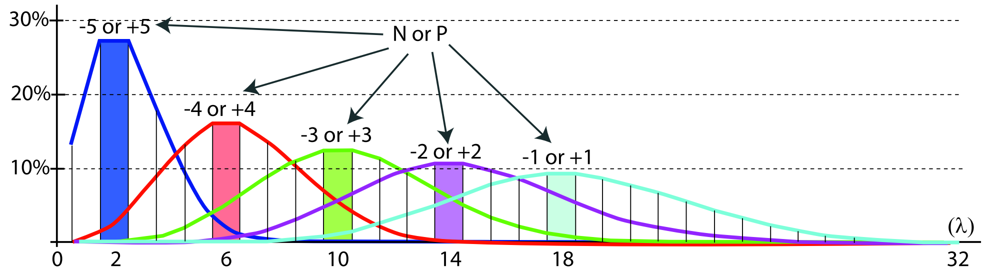


Fig. 3 Corresponding Poisson distribution used for N or P with a total 41 samplings (from 32 to 40 are almost null). Notice that for a correct approximation, the discrete range used (i.e., for λ and k) should be at least of 3 times of the natural discrete range of input data (i.e., here N, P)

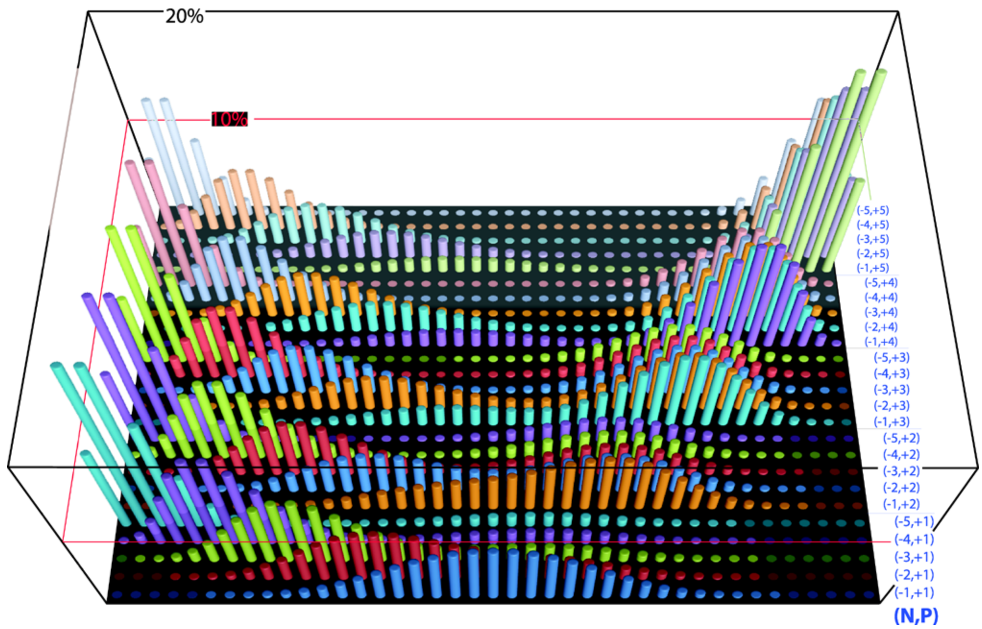


Fig. 4 Graphical representation of the sum of factors N and P: as each parameter has 5 possible values, we obtain 25 possible statistical probability cases

chances are precise. As they become small (e.g., 1), the corresponding probabilities should be weak and occur within a large range.

Note that to accelerate computations in the pipeline and therefore to allow the real-time rendering not to be interrupted by emotion computation, most Poisson distributions and weighted sums are pre-computed into tables.

5.1 Valence model

The valence parameter, i.e., the x-axis of *Russell's circumplex model of emotion*, is determined by using the four parameters (two lexical, two language) derived from the data mining model.

Lexical influence As any sentence can have both positive and negative sentiments, we have seen in the previous section that two parameters (N, P) can be extracted. In a sentence, the vocabulary also indicates how strong the emotion intensity can be interpreted. Furthermore, the intensity of those factors influences the probability of how sure the sentence tends to be positive or negative. For instance, a (N, P) couple of $(-1, +5)$ indicates that:

1. $N = -1$: there is little chance that this sentence has a negative valence, and if so, the sentence can be neutral, moderately negative, or even strongly negative;
2. $P = +5$: there is a high chance that the sentence is strongly positive.

We have to take into account both negative and positive eventualities. As N and P both have five possible values, only 25 possibilities exist. Figure 4 illustrates all possible cases and shows how intuitive and potentially efficient our model is. The first couple $(-1, +1)$ is the result of a non-intensive vocabulary. In this example, we know that from the lexical point of view, the emotion is clearly not strongly negative or positive: it can be neutral, lightly positive or negative. Hence probabilities to be at either extreme cases have to be almost null, and the medium emotion region should be kept weak but large. On the other hand, when the values of N or P are large, the probabilities to be clearly negative or positive should be strong and cover a relatively small region. Using Poisson distribution clearly demonstrates such probabilities: a large centered flat curve, almost null at extremum.

Language influence The first language parameter h being a "happiness" percentage naturally influences the weighted

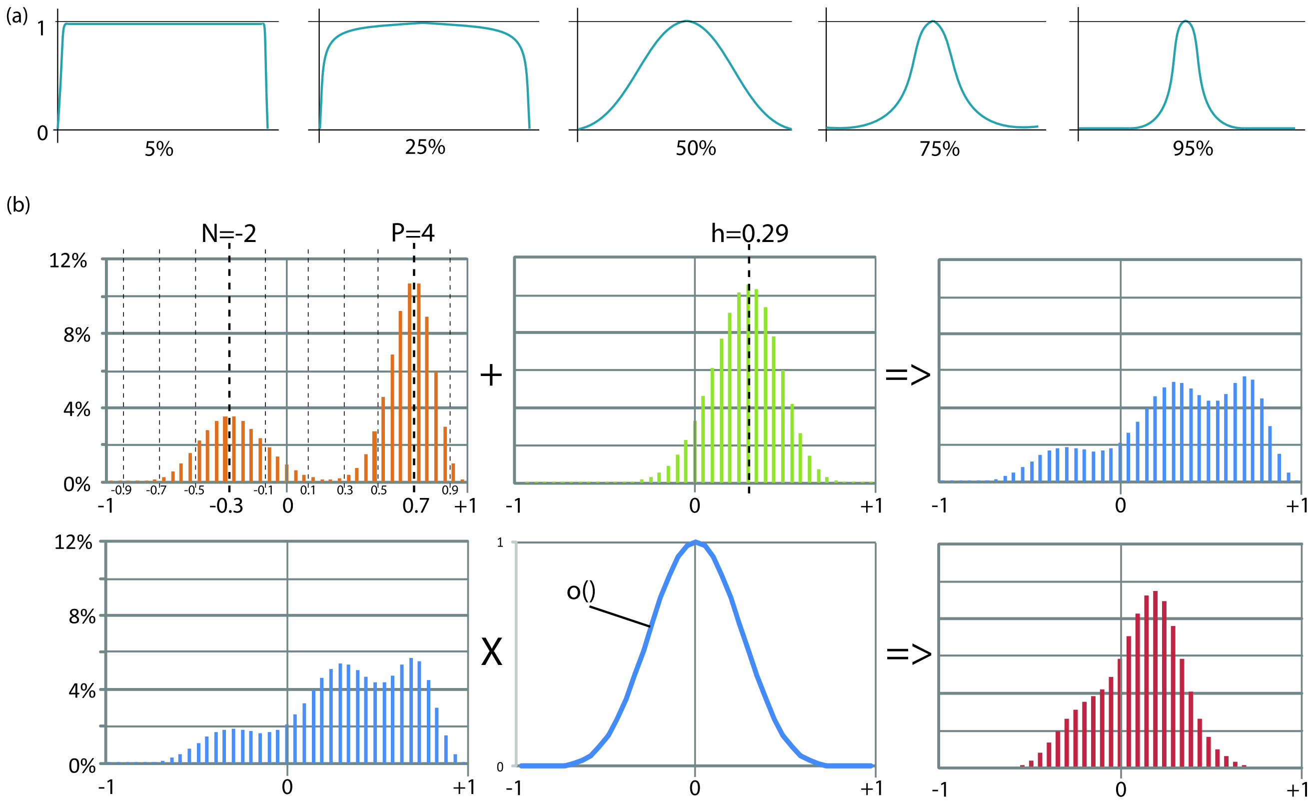


Fig. 5 From data mining values to circumplex model of emotion: (**a**) objectivity curve: the less objective the text (e.g., *top-left* scheme 5%), the more extreme the emotions; (**b**) valence computation with the language influence: in the *first row*, weighed sum of the couple $(N, P) = (-2, +4)$ (*in orange*) is multiplied by the Poisson distribution corresponding to the language "happiness" coefficient $h = 0.29$ (*in green*), assuming the number of words of the input sentence (*in blue*, also called 'c' in Fig. 9). In the *second row*, the resulting histogram (*in blue*) is then transformed by the curve $\Im$ (i.e., *blue line*) to simulate the objectivity of the text

sum from (N, P) described in the previous paragraph—see the sum presented in the first row of Fig. 5(b). Notice that the probability of h depends on the number of words used in the sentence, i.e., the larger, the better. To take into account these properties, we propose moderately or slightly increasing the happiness parameter influence with the following damping function f_h and with w_a, the average number of words in an English sentence ($w_a = 14.3$):

$$f_h = \ln(w_a h)/\big(2\ln(w_a)\big). \tag{10}$$

The objectivity o also plays an important role in the computation of the valence. Basically, the computed valence changes when the objectivity tends to zero (i.e., very subjective), and the extremum should be minimized as objectivity tends to 1 (i.e., very objective). To influence the valence, we therefore use a moderation function $\Im$, with m_s being the maximum sampling of Poisson displacement and α being any nonnull natural number:

$$\forall k \in [0, \dots, m_s] : \Im_k = \left(\frac{\cos(\frac{k\pi}{m_s}) - \pi + 1}{2}\right)^{\alpha^2 k^\alpha}. \tag{11}$$

Five examples of this equation curves and the two steps of the influence of language are presented in Fig. 5(a) and Fig. 5(b), respectively.

5.2 Arousal model

While dynamics plays an important part in getting arousal across, it is linked to other constructs such as how genuine an emotions is, or how trustworthy a person is [22]. In our approach, the arousal, i.e., the y-axis of *Russell's circumplex model of emotion*, can only be based on the intensity of the vocabulary. The arousal function $a()$ can be defined by:

$$a_{N,P} = (P - N - 2) \times 0.125. \tag{12}$$

This function $a()$ would be sufficient if and only if the occurrences of (N, P) (that have been computed based on a large database) were well balanced, i.e., followed a continuous and symmetric curve. The right graph of Fig. 6 demonstrates the irregularities of occurrences. Hence, to obtain a "well balanced" statistic, we should modify function $a()$ according to these statistics. To do so, for any percentage oc-

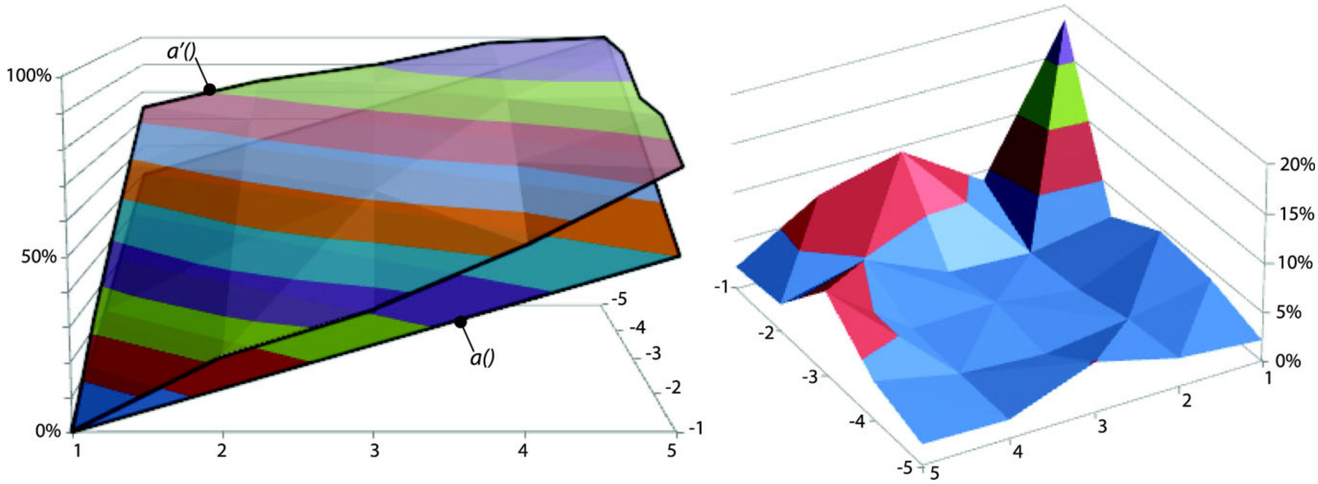


Fig. 6 Arousal functions: the ideal function $a()$ and the corresponding weighted function $a'()$ (*left*); occurrences of (N, P) based on about 4.2 million samples, 2,554,260 from BBC forums and 1,628,506 from Digg (*right*)

Table 1 Arousal function $a_{N,P}$ and corresponding weighted arousal function $a'_{N,P}$

$N\backslash P$	1	2	3	4	5	1	2	3	4	5
−1	0.0	12.5	35.4	37.5	50.0	0.0	22.2	35.4	53.5	74.7
−2	12.5	25.0	37.5	50.0	62.5	23.5	45.5	62.4	70.8	85.0
−3	25.0	37.5	50.0	62.5	75.0	40.8	62.4	74.7	85.9	86.4
−4	37.5	50.0	62.5	75.0	87.5	58.2	74.4	86.9	93.8	97.3
−5	50.0	62.5	75.0	87.5	100.0	74.9	84.8	90.3	98.6	100.0

currence $o_{N,P}$ of any couple (N, P), we define $a'()$ by

$$a'_{N,P} = a_{N,P} \times \left(1 - o'_{N,P} \times (1 - a_{N,P})\right), \tag{13}$$

with

$$o'_{N,P} = S_{\min} + \frac{S_{\max} - S_{\min}}{P_{\max} - P_{\min}} \times (P_{\min} - o_{N,P}) \tag{14}$$

and with $P_{\min}$, $P_{\max}$, $S_{\min}$, and $S_{\max}$ respectively being the minimum and maximum probability and the scaling factor of $o()$.

Arousal matrices for every possible couple (N, P) and for both $a_{N,P}$ and $a'_{N,P}$ are shown in Table 1. Notice that the use of $a'()$ remains an open issue, as some psychologists consider that people tend to express more negative than positive opinion.

6 Rendering emotion

6.1 Colored circumplex model of emotion

The purpose of this paper is to render a metaphor of emotion. Therefore, not only is there no need for advanced graphics realism, but also for realistic textures, e.g., human skin pigmentations, since it would add noise [33] to the data that should be emphasized: the correlation between textual and rendered expression. This expression can graphically be associated with geometry or color.

Whereas in the next section we present a simple model to transform the geometry of the face according to emotion, mainly focusing on the mouth, eyes, and eyebrows, this section presents the correlation between emotion and color. In the literature, even if colors are strongly dependent on culture, common trends can be established: red for passion, green for hope, and blue for depression. Figure 7 illustrates the idea of the color model [16] associated with the Russell circumplex model. To do so, we associate v and a parameters to *RGB* attributes as follows (with $-1 < (a, v) < 1$):

$$C_{R,G,B} = \frac{a+1}{2}, \frac{v+1}{2}, \frac{1-v}{2}. \tag{15}$$

6.2 Virtual human body emotion

In order to visualize the emotion of the VH, we modify the facial geometry based on the valence and arousal values v and a generated from our lexical and language classifiers and the probabilistic emotion generator. The control mechanism is based on Ekman's Facial Action Coding System

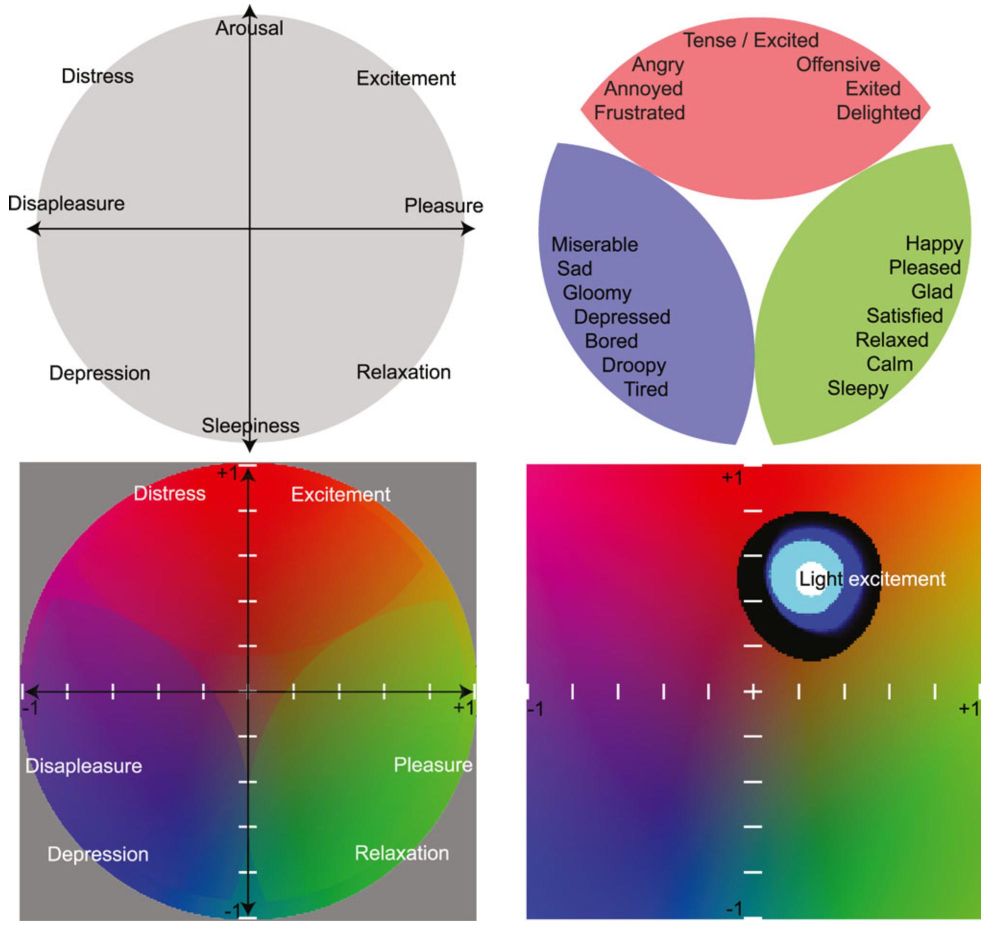


Fig. 7 From Russell's model of emotion to an association between emotion and colors with a graphical spot interpreting relevant statistical emotional region

Table 2 Facial animation based on action units with 'Xrot', 'Yrot', 'Zrot' respectively corresponding to the x-, y-, and z-axis rotations. Face part: animating face region; Action Unit(s): FACS AU(s) that correspond to the face part; Effect: axis related to circumflex model; Constraint: animating axis and *min* to *max* angle constraints

Face part	Action unit (s)	Effect (v, a)	Constraint
Head	AU 53 + 54	Arousal	$X, 5, -5$
Eye Brow *Zrot*	AU 1 + 2	Arousal	$Z, -45, 45$
Eye Brow *Xrot*	AU 4	Valence	$X, -15, 15$
Eye Lid	AU 5 + 7	Valence + Arousal	$X, 20, -10$
Cheek	AU 6	Arousal	$X, 36, -18$
Upper Lip	AU 10	Valence	$X, -15, 15$
Lower Lip	AU 16	Valence	$X, -15, 15$
Lip Corner *Zrot*	AU 12 + 15	Valence	$Z, -60, 60$
Lip Corner *Yrot*	AU 20 + 23	Arousal	$Y, -30, 30$
Jaw	AU 26	Intensity	$X, 0, 15$

(FACS) action units [13]. The summary of the geometric modifications is provided in Table 2. Only ten subsets of action unit (AU) have been selected, which are mainly influenced by the facial expression of emotions. Facial expressions of nine different emotions (extreme cases + neutral) are depicted in Fig. 8. The horizontal axis represents valence (from −1.0 to 1.0), and the vertical axis represents arousal (from −1.0 to 1.0). Applying the third axis that represents intensity (see Sect. 8) should enrich the emotional expression.

In order to augment the realism of the virtual environment, the head orientation of VH is directed towards the speaker's position. Finally, during conversation, a motion-captured talking body animation is activated.

7 Results

7.1 Graphical interpretation of emotions

Figure 9 illustrates the results of three simple cases for the interpretation of emotion found in text using our model. The first six histograms on the left (from the first left to the second row right) present histograms of the input parameters N and P, of their weighted sum, of the input parameter h, a resulting histogram of the combination of lexical and language valence c (see the blue curve of Fig. 5), and a curve $o()$ of the input parameter o. The histogram in the middle illustrates a double-histogram of the output valence v (i.e.,

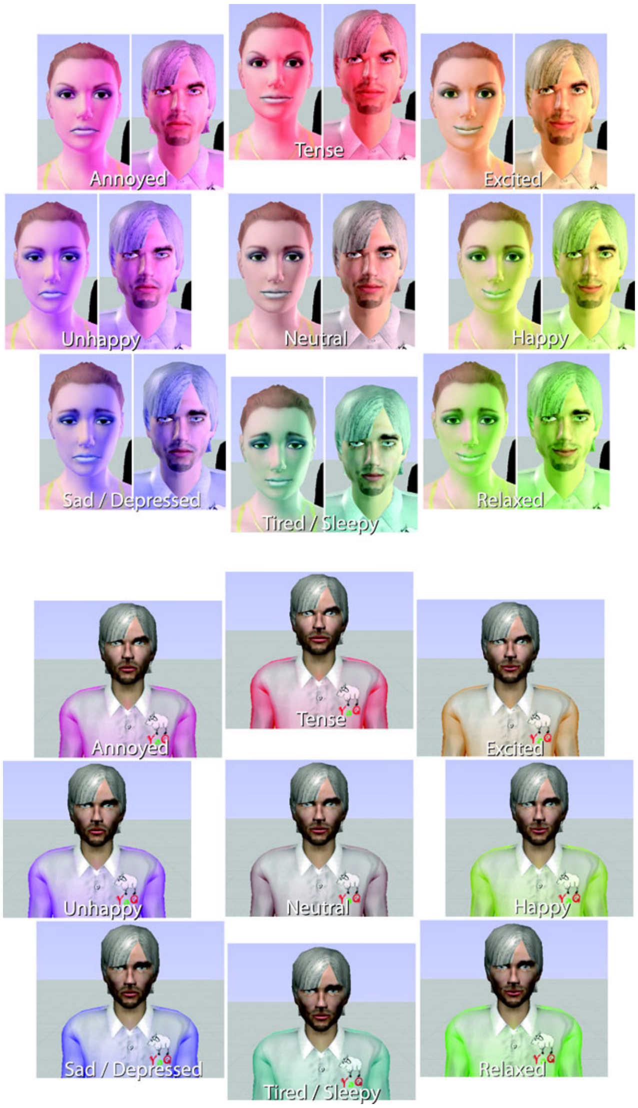


Fig. 8 *Upper image*: Facial emotion generated from valence and arousal values (v and a). The horizontal and vertical axes represent v and a, respectively, and we applied directly to VH the ambient light from the circumplex model (see Sect. 6.1). *Lower image*: Same as the *upper image* from a different point of view and without ambient light

$o(c)$) and the output arousal a (i.e., weighted function of max sum of N, P). This histogram is then mapped into the 2D colored emotion graph on the right. Notice that from the 2D resulting histogram, the final decision is made according to either manually edited rules, thresholds, or random computer choice. The whole process to compute the emotion, i.e., from the beginning to the end of the pipeline, including the rendering, takes less than 10 ms.

Common cases Based on the following real sentences sampled from Digg, each example of Fig. 9 shows very different emotional states.

Row (a)—"179 *Diggs in* 30 *seconds, I am a Digg Jedi Knight. Not too shabby for using a trackball that is over* 5 *years old and wearing out. LoL*" $\Longrightarrow (-5, +4)$; $[0.90, 0.87]$. This text uses a strong and energetic vocabulary with a language analysis definitively positive and objective. The diagram proposes with a high accuracy (i.e., a small spot) a positive and dynamic attitude (potentially aggressive but not angry);

Row (b)—"*I like the part where the guy points out the protesters' hypocrisy*" $\Longrightarrow (-2, +3)$; $[0.22, 0.14]$. This sentence corresponds to a weak negative and moderately positive lexical sentence with a fairly negative and subjective language. The resulting diagram shows naturally a weak negative valence and arousal, demonstrating a possible pessimistic or sarcastic emotion—notice how big the spot is as it is difficult to judge the intensity of this emotion;

Row (c)—"*Interesting and useful info thnx, bookmarked!*" $\Longrightarrow (-1, +3)$; $[0.90, 0.87]$. The text is clearly positive in terms of vocabulary and language. Furthermore, this sentence is objective, which results in a happy and potentially satisfied emotion.

7.2 Virtual human emotion rendering

Our main demonstration, including life testing, dialogs, and pipeline explanation, can be found at: http://fromsentencetoemotion.serveftp.org.

We have defined and implemented an xml format for each dialog scene. The xml file is composed of three main fields: dialog info, characters, and dialog.

Dialog info The dialog info field is the header of the textual conversation. It consists of the number of avatars (participants), the subject and the source of dialog, i.e., language, dialog type, title, authors, edition, and year.

Characters The character field describes each avatar's attributes such as the actor's id, name, gender, age, role, and general mood.

Dialog The dialog field is the actual conversation between avatars. Each dialog consists of speaking time, speaking actor id, addressee id, speaking sentence, speaking speed, emotion, emotional valence, and emotional arousal.

By using this xml, we can also pre-define the mood and emotion of each character on each dialog sentence. By parsing the xml, our animation engine automatically constructs a conversational scene with avatars and agents. As illustrated in Fig. 1, the VH inner circle represents avatars, and the VH outer circle represents surrounding agents for crowd emotion.

In order to understand the emotion of each dialog, the communicated text is presented on the top left of the screen. The color that corresponds to the emotion of the

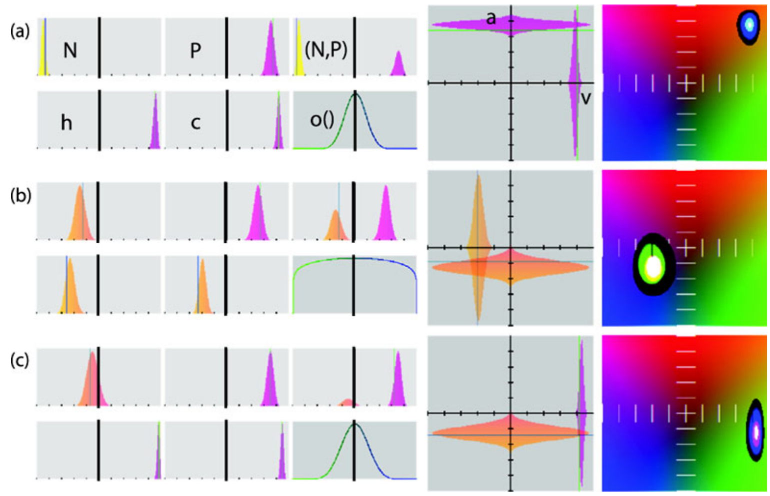


Fig. 9 Three examples of typical cases of probabilistic emotional histograms

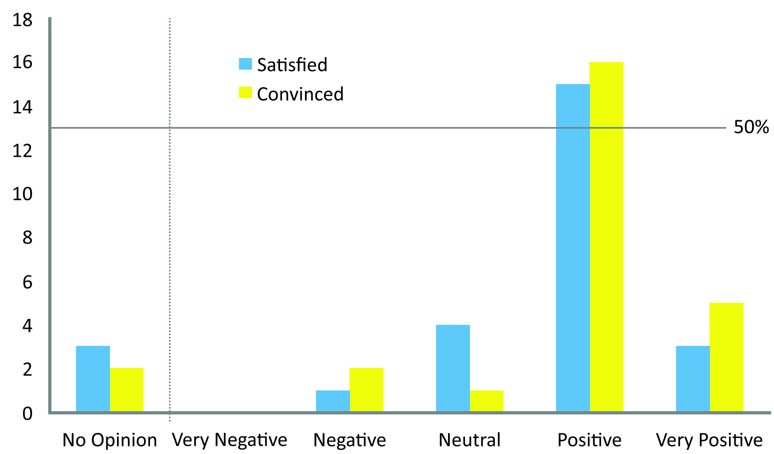


Fig. 10 Results of a survey with 26 persons that have tested our system: *in blue*, satisfaction after testing our *graphical metaphor of emotion*; *in red*, whether such an idea can be useful in virtual reality

current sentence is applied both to the text and to the avatar's body. We can adjust the ambient variables applied to the speaking avatar. The analyzed and synthesized emotional status is also displayed on the top of the head of each avatar. The facial expressions are generated based on these values. Lastly, for the entire scene, the heads of the avatars and agents are directed towards the current speaker.

An experiment with 26 persons (3 women and 23 men from 27 to 55 years old) was conducted to justify our model. We asked each of them to test our system by writing sentences *without thinking of a specific feeling relative to a personal experience*. Then, we asked whether they were satisfied by the result and if they were convinced by the usefulness. Results are shown in Fig. 10.

To test our system on different types of media, we also applied a comic and movie script to our proposed system, and compared the emotional difference between the real and virtual scene. The results are depicted in Fig. 11.

8 Conclusion

This paper has presented a model to generate a graphical metaphor of text emotion applied to 3D VH face expressions. We first presented our original processes pipeline including the following three procedural steps: an emotional statistical model and a data mining model based on a very large database (4.2 million samples); a two dimensional model based on the Poisson distribution to derive the *valence* and the *arousal* parameters to give a statistically good correlation with *James Russell's circumplex model of emotion*; and last, a rendering model including colors and facial emotions. To show the flexibility and efficiency of our model, we described emotion graphics metaphors resulting from real sentences (taken from Internet fora). Finally, we demonstrated the capabilities of our graphics metaphor through simulating dialog with real-time VH rendering.

We are currently working on a more realistic model suitable for any conversation. Text communication such as prewritten dialog (e.g., a theater play) can be interpreted in dif-

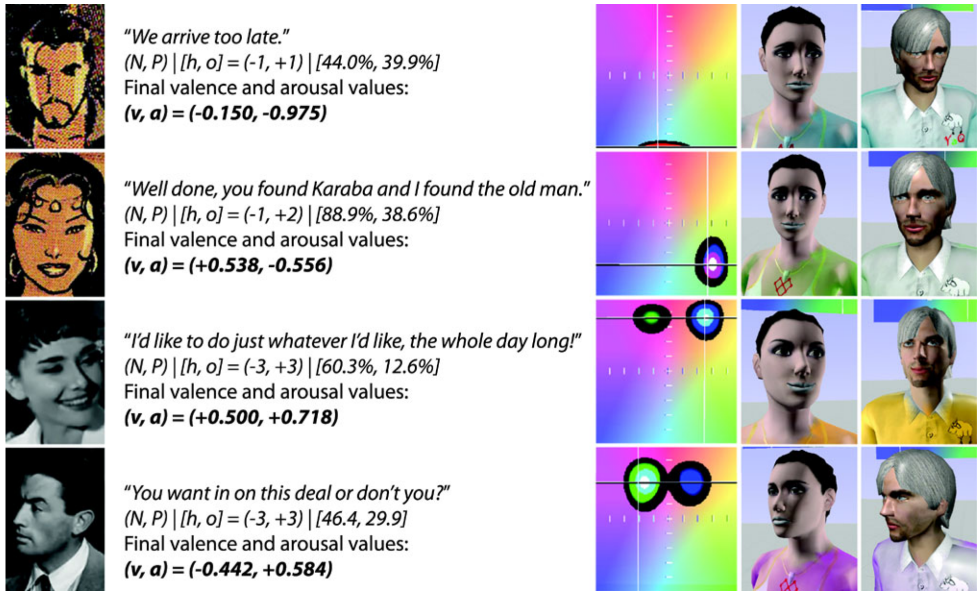


Fig. 11 Results of emotion generation from the script of comic and movie. Comic: E. Marini, S. Desberg, *The Scorpion*, Dargaud, 1993; Movie: W. Wyler, [A. Hepburn, G. Peck], *Roman Holiday*, 1953

ferent ways. Three main factors affect/alter/modify the direct emotion written in the text: the characteristic/orientation of the actors, the historical background (long term memory or short term event), and the social context. Furthermore, as Fontaine et al. explained, the world of emotions is certainly not two-dimensional [17]. We have shown in this paper that the presented model could not directly simulate the third dimension called (*potency* or *intensity*). Including at least three dimensions for the representation of emotion is certainly the most challenging perspective of this work.

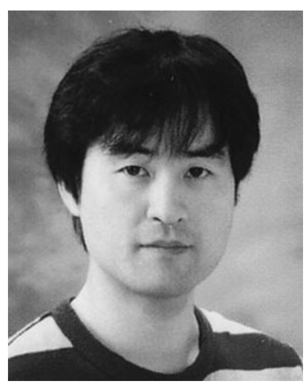
Acknowledgements The authors wish to thank Dr. Arvid Kappas for his advices concerning our basic model of emotion, as well as Mr. Quentin Silvestre for his technical help. Our thoughts also go to Dr. Barbara Yersin and Dr. Jonathan Maim, the original CS developer of the VH animation engine called YaQ. This work was supported by two grants: the European Union CYBER**EMOTIONS** and the SNF AERIAL**CROWDS** projects.

References

1. Allbeck, J., Badler, N.: Toward representing agent behaviors modified by personality and emotion. In: AAMAS 2002, Workshop on Embodied Conversational Agents—Let's Specify and Evaluate Them!, Bologna, Italy (2002)
2. Amaya, K., Bruderlin, A., Calvert, T.: Emotion from motion. In: GI '96, pp. 222–229. Canadian Information Processing Society, Ontario (1996)
3. Alm, C.O., Roth, D., Sproat, R.: Emotions from text: Machine learning for text-based emotion prediction. In: Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, pp. 579–586, Vancouver, Canada, October 2005
4. Badler, N.I., Allbeck, J., Zhao, L., Byun, M.: Representing and parameterizing agent behaviors. In: CA '02: Proceedings of the Computer Animation, pp. 133–143. IEEE Computer Society, Washington (2002). ISBN:0-7695-1594-0
5. Becheiraz, P., Thalmann, D.: A model of nonverbal communication and interpersonal relationship between virtual actors. In: CA '96, p. 58, Washington, DC (1996)
6. Cassell, J.: Embodied Conversational Agents. MIT Press, Cambridge (2000)
7. Cassell, J., Högni Vilhjálmsson, H., Bickmore, T.: BEAT: The behavior expression animation toolkit. In: SIGGRAPH 2001, Annual Conference Series, pp. 477–486. ACM Press, New York (2001)
8. 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, pp. 413–420. ACM, New York (1994)
9. Coyne, R., Sproat, R.: Wordseye: an automatic text-to-scene conversion system. In: SIGGRAPH, pp. 487–496 (2001)
10. Chittaro, L., Serra, M.: Behavioral programming of autonomous characters based on probabilistic automata and personality. J. Comput. Animat. Virtual Worlds **15**(3–4), 319–326 (2004)
11. Costa, M., Zhao, L., Chi, D.M., Badler, N.I.: The EMOTE Model for Effort and Shape, pp. 173–182. ACM Press, New York (2000)
12. Davidson, R.J., Scherer, K.R., Goldsmith, H.H.: Handbook of Affective Sciences. Oxford University Press, London (2003)
13. Ekman, P., Friesen, W.V.: Facial Action Coding System. Consulting Psychologists' Press, Palo Alto (1978)
14. Egges, A., Molet, T., Magnenat-Thalmann, N.: Personalised real-time idle motion synthesis. In: Computer Graphics and Applications, pp. 121–130. IEEE Computer Society, Los Alamitos (2004)
15. Feldman-Barrett, L., Russell, J.A.: The structure of current affect. Curr. Dir. Psychol. Sci. **18**(1), 10–14 (1999)
16. Fagerberg, P., Ståhl, A., Höök, K.: Emoto: emotionally engaging interaction. Personal Ubiquitous Comput. **8**(5), 377–381 (2004)
17. Fontaine, J.R., Scherer, K.R., Roesch, E.B., Ellsworth, P.: The world of emotions is not two-dimensional. Psychol. Sci. **18**(12), 1050–1057 (2007)
18. Grillon, H., Thalmann, D.: Simulating gaze attention behaviors for crowds. Comput. Animat. Virtual Worlds **20**(2–3), 111–119 (2009)
19. Hess, U., Adams, B.R., Grammer, K., KLeck, R.E.: Face gender and emotion expression: Are angry women more like men? Vision **9**(12), 1–8 (2009)

20. Hatzivassiloglou, V., Wiebe, J.: Effects of adjective orientation and gradability on sentence subjectivity. In: Proceedings of the 18th Conference on Computational Linguistics, 2000, Saarbrücken, Germany, pp. 299–305. Association for Computational Linguistics, Morristown (2000)
21. Kappas, A.: What facial activity can and cannot tell us about emotions. Behav. Process. **60**, 85–98 (2003)
22. Krumhuber, E., Manstead, A., Cosker, D., Marshall, D., Rosin, P., Kappas, A.: Facial dynamics as indicators of trustworthiness and cooperative behavior. Emotion **7**(4), 730–735 (2007)
23. Kshirsagar, S., Magnenat-Thalmann, N.: A multilayer personality model. In: SMARTGRAPH '02: Proceedings of the 2nd International Symposium on Smart Graphics, pp. 107–115. ACM, New York (2002)
24. Maïano, C., Gobron, S., Pergandi, J.-M., Therme, P., Mestre, D.: Affective and behavioral responses to a virtual fear elicitation scenario. In: VRIC'07 (Laval Virtual 07), pp. 83–91, April 18–20 2007
25. Masuko, S., Hoshino, J.: Head-eye animation corresponding to a conversation for cg characters. In: EG'07, pp. 303–311 (2007)
26. Macdonald, C., Ounis, I., Soboroff, I.: Overview of the trec-2008 blog track. In: TREC 2008 (2008)
27. Manning, C.D., Schuetze, H.: Foundations of Statistical Natural Language Processing, 1st edn. MIT Press, Cambridge (1999)
28. Musse, S.R., Thalmann, D.: Hierarchical model for real time simulation of virtual human crowds. IEEE Trans. Vis. Comput. Graph. **7**(2), 152–164 (2001)
29. Pelachaud, C.: Modelling multimodal expression of emotion in a virtual agent. Philos. Trans. R. Soc. B, Biol. Sci. **364**(1535), 3539–3548 (2009)
30. Pelachaud, C.: Studies on gesture expressivity for a virtual agent. Speech Commun. **51**(7), 630–639 (2009)
31. Perlin, K., Goldberg, A.: Improv: a system for scripting interactive actors in virtual worlds. In: SIGGRAPH'96, pp. 205–216 (1996)
32. Picard, R.: Affective Computing. MIT Press, Cambridge (1998)
33. Park, K.H., Kim, T.Y.: Facial color adaptive technique based on the theory of emotion-color association and analysis of animation. In: MMSP, pp. 861–866. IEEE Signal Processing Society (2008)
34. Pang, B., Lee, L.: Opinion Mining and Sentiment Analysis. Now Publishers Inc., Boston (2008)
35. Pfeiffer, P.E., Schum, D.A.: Introduction to Applied Probability. Academic Press, New York (1973)
36. Russell, J.A., Bachorowski, J., Fernandez-Dols, J.M.: Facial and vocal expressions of emotion. Annu. Rev. Phychol. **54**, 329–349 (2003)
37. Ruttkay, Z., Noot, H., ten Hagen, P.: Emotion disc and emotion squares: tools to explore the facial expression space. Comput. Graph. Forum **22**(1), 49–53 (2003)
38. Russell, J.A.: A circumplex model of affect. J. Pers. Soc. Psychol. **39**, 1161–1178 (1980)
39. Stone, M., DeCarlo, D., Oh, I., Rodriguez, C., Stere, A., Lees, A., Bregler, C.: Speaking with hands. ACM Trans. Graph. **23**(3), 506–513 (2004)
40. Sebastiani, F.: Machine learning in automated text categorization. ACM Comput. Surv. **34**(1), 1–47 (2002)
41. Shao, W., Terzopoulos, D.: Autonomous pedestrians. Graph. Models **69**(5–6), 246–274 (2007)
42. Su, W.P., Pham, B., Wardhani, A.: Personality and emotion-based high-level control of affective story characters. IEEE Trans. Vis. Comput. Graph. **13**(2), 281–293 (2007)
43. Unuma, M., Anjyo, K., Takeuchi, R.: Fourier principles for emotion-based human figure animation. In: SIGGRAPH'95, pp. 91–96. ACM, New York (1995)
44. Velasquez, J.D.: Modeling emotions and other motivations in synthetic agents. In: Proceedings of AAAI97 (1997)
45. Wiebe, J.: Learning subjective adjectives from corpora. In: IAAI'00, pp. 735–740. AAAI/MIT Press, Cambridge (2000)
46. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis. In: HLT '05, pp. 347–354. ACL, Morristown (2005)
47. Yu, Q., Terzopoulos, D.: A decision network framework for the behavioral animation of virtual humans. In: SCA '07, pp. 119–128. EG Association, Aire-la-Ville (2007)
48. Zhe, X., Boucouvalas, A.: Text-to-emotion engine for real time internet communication. In: International Symposium on Communication Systems, Networks and DSPs, pp. 164–168 (2002)

Stephane Gobron since the end of his PhD in Japan (2001), he has mainly been doing research in Computer Graphics where his research interests primarily focus on geometrical modeling and virtual reality. In France, he has been teaching computer science courses where he has been in charge of the Computer Graphics Bachelor department. He is currently a member of the VRLab at EPLF, Switzerland, managing a small team, focusing the R&D on 3D visualization of virtual human, and participating in teaching at the graduate level.



Junghyun Ahn is a Post-Doc Researcher at EPFL VRLab. He received his BS, MS, and PhD degrees in Computer Science from KAIST in 1998, 2000, and 2007, respectively. He was a Senior Researcher in KIEE and Short-term Visiting Scholar in UPenn. His research interests are crowd animation, motion capture, virtual reality, and cognitive modeling.

Georgios Paltoglou is a post-doctoral researcher at the University of Wolverhampton, being part of the Statistical Cybermetrics Research Group. His work is focused on opinion mining and sentiment analysis. His PhD was on Information Retrieval with a focus on Distribution Information Retrieval systems. He received his PhD from the University of Macedonia, Greece.

Michael Thelwall is a Professor of Information Science and leader of the Statistical Cybermetrics Research Group at the University of Wolverhampton, UK. He is also a visiting fellow of the Amsterdam Virtual Knowledge Studio, a Docent at Åbo Akademi University Department of Information Studies, and a research associate at the Oxford Internet Institute. He has published 152 refereed journal articles, seven book chapters, and two books including Introduction to Webometrics. Mike Thelwall is an associate editor of the Journal of the American Society for Information Science and Technology and sits on eight other editorial boards.

Daniel Thalmann is a Director of EPFL VRlab, Switzerland. He is a pioneer in Virtual Humans. He is a coeditor-in-chief of the Journal of Computer Animation and Virtual Worlds and editor or several other journals. He has published numerous papers in Graphics, Animation, and VR. He is coauthor of several books, including Crowd Simulation (2007). He received his PhD in Computer Science from University of Geneva and an Honorary Doctorate from University Paul-Sabatier in Toulouse, France.