

Constructing Knowledge for Automated Text-Based Emotion Expressions*

Siska Fitrianie and Leon J.M. Rothkrantz

Abstract: *Humans are used to convey their thought through their (conscious or unconscious) choice of words. Some words possess emotive meaning together with their descriptive meaning. We develop a prototype of a synthetic 3D face that shows emotion associated to text-based speech in an automated way. As a first step, we studied how humans express emotions in face to face communication. Based on this study, we develop a 2D affective lexicon database and a set of rules that describes dependencies between linguistic contents and emotions. The result described in this paper proposes an initial step for developing knowledge for an affective-based multimodal fission.*

Key words: *Emotion, multimodal communication, knowledge acquisition, natural language processing.*

INTRODUCTION

As described by many theorist (e.g. [1][2][8]), emotion expressions have three major functions: (1) they contribute to the activation and regulation of emotion experiences; (2) they communicate something about internal states and intentions to others; and (3) they activate emotions in others, a process that can help account for empathy and altruistic behavior. These expressions could be shown in the words we used in speech and our nonverbal behaviors. The challenge is that nonverbal behaviors do not occur randomly, but rather are synchronized to one's own speech or to the speech of other ([3][5][17]). Seeing faces, interpreting their expression, understanding the linguistic contents of speech are all part of human communication.

Most existing human computer interaction systems, to the best of our knowledge, have not considered explicitly the emotions existing in the speech content. The difficulty lies in the fact that emotional linguistic content consists of entities of complexity and ambiguity such as syntax, semantics and emotions. The use of simple templates has proven to be useful for the detection of subjective sentences and of words having affective semantic orientation (e.g. [9][13]). These simplistic models describe how words with an affective meaning are being used within a sentence, but fail to offer a more general approach. Furthermore, current developments of affective lexicon database (e.g. [16][19]) are based on subjective meaning of the emotion words and do not provide information about the (relative) distance between words in regards to their emotion loading context. The lack of a large-scale affective lexicon resource database makes a thorough analysis difficult. As a consequence, although important, an automated emotional expression from natural language is still rarely developed.

At MMI-Group TUDelft, there is a project running on natural human computer interaction. We have developed a system that allows average users to generate facial animations using a syntactic 3D face based on Facial Action Coding System (FACS – [5]) in a simple manner [20]. We also have built a dictionary of facial expressions (FED – [11]) that stores the facial expressions that naturally occur in face to face communication. The project described in this paper aims at developing a system that is able to reason emotions automatically from natural language text and show appropriate facial expressions as its response to the emotion content of the text. In the following sections we will give an overview of the system. Further we will concentrate on developing an affective lexicon database and the reasoning for emotions analysis from text as well as an experiment to generate the knowledge. The result described in this paper proposes an initial step for developing knowledge for an affective-based multimodal fission.

* The research reported here is part of the Interactive Collaborative Information Systems (ICIS) project, supported by the Dutch Ministry of Economic Affairs, grant nr: BSIK03024

SYSTEM OVERVIEW

Fig. 1 shows the pipeline processes of the developed system. The Emotion Analysis module has a parser that associates input text to emotions and transforms it into XML format. For this purpose, it exploits the affective lexicon database. The Reaction Module processes the XML results by assigning appropriate facial expressions. It has a FED in which a lexeme contains an emblem of a facial expression, a description of which AUs are activated, semantic interpretations, and an example using a synthetic face [11].

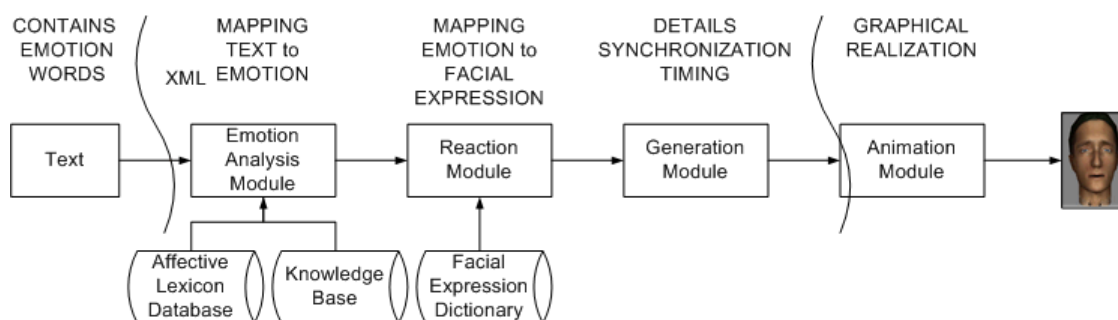


Figure 1 System architecture

The Generation module plans the display to be synchronized with speech. It estimates of speech timings and constructs an animation schedule prior to execution. The Animation module uses the results to generate facial animation based on “facial script language”, where basic variables are AUs and their intensity. It employs a parametric model for facial animation and a method for adapting it to a specific person based on performance measurements of facial movements [20].

DATA ACQUISITION EXPERIMENT

The purpose of our experiment was to find out: (1) what kind of emotion expressions are shown most often in a conversation; (2) what is the typical course of particular expressions; (3) whether the expressions depends on each other; and (4) whether the expressions are related in any way to used certain words. To collect data for the analysis we prepared ten scenarios of dialogs between two characters with diverse situations which evoke various affective states and asked ten participants to perform a role of one character as many expressions as possible from the scenarios. To obtain diversity of emotion expressions, provided scenarios contained a high number of punctuation marks. The facial expressions of participants during the experiment were recorded on a video recorder to be analyzed afterward. The video sequences were converted and stored as MPEG-2 stream. They were sampled at 25 frames per second and saved with 645 KB/sec bit rate. As the first step, three independent observers marked the onset and offset of an expression. In the next step, these expressions were labeled according to the context. In the final step, we also collected emotion words used in each expression. The agreement rates between the observers in both steps were about 73%.

The experimental results indicated that our participants showed most of the time a neutral face. However, we managed to capture in total 40 different facial expressions; about 20-35 different expressions per participant in each dialog, and 119 emotion words. Our experimental results were endorsed by an experiment conducted by [4]. He found 41 displayed emotion expressions actually used to appraise a product (see table 3).

To analyze distribution of the elapsed time of facial expressions, we plotted appropriate histograms for each emotion label with interval length of 5 frames and the range from 0 – 104 frames (all expression persisting longer than 4 seconds where put into the last interval). From the histograms, we noticed that emotion expressions mostly appear for a rather short period of time, somewhere between half and just more that one second

(10-30 frame). However, 40% of “surprise” expression could appear until 4 second. Expressions: “anger” and “disgust” had a shape of distribution similar to the shape of distribution of “surprise”, but last usually a little longer than “surprise” (15-19 frames, while most “surprise” 5-9 frames). 90% of “happiness” were distributed in the range between 10-50 frames. Its histogram also contained a tail that comprises of the longest observed expression duration. We attributed this to the dual role of the expression as both short communication signal and a long-lived mood indicator.

To study the relationship between emotion expressions, we examined which facial expressions could occur at the same time. Generally, almost 13% of all segments covered another segment for at least one frame. We found a high number of occurrences of two kinds of combinations: “surprise” and “sadness” and “anger” with “disgust”. Usually “surprise” preceded and overlapped “sadness” and 65% cases of “anger” had segments that contained the whole segment of “disgust”. 50% of “disgust” were entirely enclosed in longer segments of “anger”.

To study the relationships between facial expressions and text, we determined the timing of occurrences of each word and punctuation mark by partitioning the dialog (manually) into basic constituents, which are usually formed by a single sentence and for each constituent into components, i.e. single words and punctuation mark. Then, for each component, we determined the time of its occurrence (number of frames in which the given word is pronounced). The next, we determined which components coincide with the shown expressions. This means, for each selected facial expression, we had to determine the text that starts with the component synchronized with the first frame of a given facial expression and ends with the component synchronized with the last frame of this expression. The experimental result showed that most of the facial expressions (around 63%) corresponded to the text spoken. For sentences with questions or exclamation marks, distinguishably, “surprise” is the most common facial expression displayed during a question. It appeared almost exclusively in short and single-word question, e.g. “really?”, “sure?”. We noticed that the sentences ending with exclamation mark were usually accompanied by expression “anger”.

In the final experiment, we focused on mapping the shown expressions to emotional words defined by [4]. The distance of a given facial expression from a particular emotional word was defined as the number of frames with the neutral face, which appear between the facial expressions. The results showed that 54.6% of the emotion words spoken by the participants linked to facial expressions. To check whether the participants really displayed more facial expressions for emotional words than for any other words we compared results from above with the analogous statistics for non-emotional words. The comparison showed that the use of emotion words, indeed, evoked emotions which were expressed by facial expressions (see table 1). Although, with this experiment, we still could not draw direct link between the emotion words with the facial expressions. It is because some words only occurred once or twice and some other words in different context were related to different facial expressions.

Table 1 Statistics of emotion and non-emotion words in the input text linked to facial expression

Words	Total	Linked to Expression
Non-emotion words	2206	1022 (46.3%)
Emotion words	119	65 (54.6%)

AFFECTIVE LEXICON DATABASE

Russell [18] and Desmet [4] analyzed qualitative values of emotion words based on humans’ social value and depicted them in a 2D space of pleasure and activation. The dimension can categorize emotions in a comprehensible way; however, the approach is not yet sufficient to differentiate between emotions, e.g. anger and fear fall close together on the circumplex. Kamps and Marx [12] proposed the quantitative differences

between the relatively object notions of lexical meaning and more subjective notions of emotive meaning by exploiting WordNet [7]: (1) the smallest number of synonymy (synset) steps between two words, e.g. MPL(good, bad) 4 {good, sound, heavy, big, bad}, and (2) the relative distance of a word to two reference words, e.g. EVA*(proper, good, bad) = 1 and POT*(amazed, active, passive) = 0.75. The EVA* allows us to differentiate between words that are predominantly used for expressing positive emotion (close to 1), for expressing negative emotion (close to -1), or for non-affective words (value = 0).

We used the results of the experiment above and selected the stem of 140 emotion labels and words that were found in the experiment as initial records of our database. We aimed to depict them in a 2D space, by two approaches. Firstly, based on [12], the direction and distance of a vector in the bipolar dimension represent the quality (Pleasant-Unpleasant) and intensity (Active-Passive). Fig. 2(a) shows an example of the plotting. The degree of correctness by this approach was 78%.

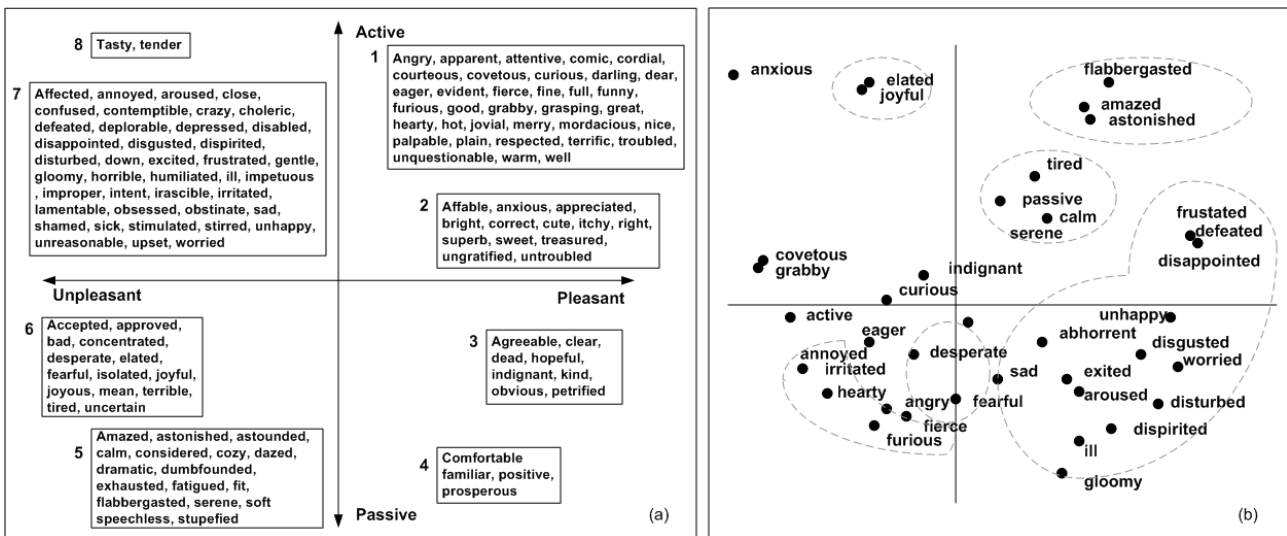


Figure 2 (a) a 2D space emotion words in octants-related Pleasantness-Activation and (b) MDS-Mapping emotion words

Finally, we applied multidimensional scaling (MDS) to represent emotion words in 2D space. We employed [12]’s MPL to construct an NxN matrix as the input. The Euclidian distances among all pairs of points were applied to measure natural distances of those points in the space. Using “similarity” (corresponding meaning) between emotion words, this procedure found the clusters that approximate the observed distance in the best way. By lowering the degree of corresponding between the Euclidian distance among points and the input matrix, the best corresponding MDS map can be achieved. Fig. 3(b) shows an example of emotion words after MDS mapping. The degree of correctness by this approach was 65%. For both approaches, manual checking was still necessary for all mistaken classified emotion words.

MAPPING TEXT TO EMOTIONS KNOWLEDGE BASE

Based on the results of the experiment above, we distinguished five types of emotional intentions: (a) emotionally active toward an object, e.g. “I hate vegetable”; emotionally directed by an object, e.g. “she treats me badly”; (c) emotions that provoked by an object, e.g. “his attitude makes me angry”; (d) emotions that experienced towards an object, e.g. “it is beautiful picture”; and (e) appraisal toward an object, e.g. “her mother is ill”. These findings were supported by Mulder et.al. [14] that stated that emotions in language as having a positive or negative orientation, an intensity and a direction toward an object or event.

We develop heuristic rules to assess an emotion intention of a sentence and associate it with an emotion type. The rules describe the subjective notion of the sentence

using three attributes: the experiencer, the attitude, and the object. The experiencer is the person in a private state of the kind attitude towards the object. By decomposing the syntax structure of the sentence and its thematic roles, we can extract these attributes including direction of the attitude toward the object (passive, active and indirect). Table 2 shows an example of the rules. Using the affective lexicon database, we can find the attitude's orientation (positive, negative or neutral), its intensity, and its classification and relative distance with other attitudes (or emotion words).

Table 2 An example of constructing rule base for extracting emotion intentions in text

Example of input:	"his attitude makes me angry"
Syntax:	subject + verb + object (noun) + (verb) + (pronoun-object + adverb)
Thematic roles:	material + verb + product + quality
Rule:	If exist(experiencer, product) and exist(object, material) and exist(experiencer, "main-role") Then attitude(experiencer="main-role", quality, product, material, direction="passive")

To control the intensity of emotion expressions from statement to another statement, we design six universal "emotion thermometers" based Ekman's universal emotion types [6]. Although we aim at using 41 emotion expressions, to simplify, we correspond the emotions with the universal emotion types (table 3). If a sentence is analyzed to have an emotion loading i , the value of all thermometers T will be calculated using equation:

$$T_i(t) = T_i(t-1) + I_i \cdot s$$

$$\forall j \neq i | T_j(t) = T_j(t-1) - d[j, i]$$

Where, i is the active universal emotion type, s is a summation factor, I is the emotion's activation degree, and j is a range of {happiness, sadness, anger, surprise, disgust, and fear}. The distance between two universal emotions $d[j, i]$ follows the work of [10]. The highest value of the thermometers is considered as the sentence's current mood. Using table 3, the emotion's intensity is the thermometer's value of the corresponding emotion type. A rule-based approach is used to govern the calculation, e.g. to detect negation, the contrast coordinate words (e.g. "but") and contrast subordinate words (e.g. "although").

Table 3 Universal emotions [6] - 41 emotion expressions [4]

Universal Emotions	Emotion Expressions
Happy	Inspired, desiring, loving, fascinated, amused, admiring, sociable, yearning, joyful, satisfied, softened.
Sad	Disappointed, contempt, jealous, dissatisfied, disturbed, flabbergasted, cynical, bored, sad, isolated, melancholic, sighing.
Surprise	Pleasantly surprise, amazed, astonished.
Disgust	Disgusted, greedy.
Anger	Irritated, indignant, hostile.
Fear	Unpleasantly surprised, frustrated, alarmed.
Neutral	Curious, avaricious, stimulated, concentrated, eager, awaiting, deferent.

CONCLUSION AND FUTURE WORK

A method for analyzing emotion loadings in text has been investigated. A rule-based approach has been selected for emotion reasoning. This gives opportunities for us to extend both our developed human computer interaction system's believability and behavior. To support the reasoning engine, we also develop an affective lexicon database, which depicted in a 2D space. The database gives information about the positive/negative orientation of a word, its intensity and its relative distance to other (emotion) words.

Future work is necessary to evaluate the knowledge applied by the developed system. Our approach described in this paper assumed that the emotion loadings in linguistic contents are explicit emotions, shown by emotion words. Our experimental results indicated that an emotion expression for a given emotion word depends mostly on the context and a given word used in various situations can have different meaning. The next step will include emotion analysis from discourse information, e.g. moods, personality characteristics, anaphoric information, background contexts of dialogues.

REFERENCES

- [1] Clore G., Cognitive Phenomenology: Feelings and the Construction of Judgment, In: L. Martin and A. Tesser (Eds.), *The Construction of Social Judgments*, NY: Lawrence Erlbaum Associates, 1992. 133-163.
- [2] Clore G., A. Ortony, Cognition in Emotion: Always, Sometimes, or Never? In: R. Lane and L. Nadel (Eds.), *Cognitive Neuroscience of Emotion*, NY: Oxford University Press, 2000, 24-61.
- [3] Condon W., W. Osgton, Speech and Body Motion Synchrony of the Speaker-Hearer. In D. Horton & J. Jenkins (Eds), *The perception of Language*, Academic Press, 1971, 150–184.
- [4] Desmet P., *Designing Emotion*, Doctoral Dissertation, Delft University of Technology, 2002.
- [5] Ekman P., About brows: Emotional and Conversational Signals, In M. von Cranach, K. Foppa, W. Lepenies, D. Ploog (Eds), *Human Ethology: Claims and Limits of a New Discipline: Contributions to the Colloquium*,. Cambridge Univ. Press, 1979, 169-248.
- [6] Ekman P., W.F. Friesen, *Unmasking the Face*. Englewood Cliffs, New Jersey:Prentice-Hall, 1975.
- [7] Fellbaum C., *WordNet: An Electronic Lexical Database*, The MIT Press, 1998.
- [8] Frijda N.H., *The Emotions*, Cambridge University Press, Cambridge, 1986.
- [9] Hatzivassiloglou V., K.R. McKeown, Predicting the Semantic Orientation of Adjectives, in *Proc. of ACL'97, Spain*, 1997.
- [10] Hendrix J., Zs.M. Ruttkay, Exploring the Space of Emotional Faces of Subjects without Acting Experience, *ACM Computing Classification System*, H.5.2, I.5.3, J.4, 1998.
- [11] de Jongh E., L.J.M. Rothkrantz, FED: an Online Facial Expression Dictionary, in *Euromedia 2004*, Eurosis, Ghent, 2004.
- [12] Kamps J., M. Marx, Words with Attitude, in *Proc. of Global WordNet CIIL'02*, India, 2002, 332-341.
- [13] Liu H., H. Lieberman, T. Selker, A Model of Textual Affect Sensing using Real World Knowledge, Technical Report, MIT Media Laboratory, Cambridge, 2003.
- [14] Mulder M., A. Nijholt, M. den Uyl, P. Terpstra, A Lexical Grammatical Implementation of Affect. In *TSD'04*, 171-178, 2004
- [15] Ortony A., G. Clore, A. Collins, *the Cognitive Structure of Emotions*, Cambridge University Press, 1988.
- [16] Ortony A., G. Clore, M. Foss, the Referential Structure of the Affective Lexicon, *Cognitive Science*, 1987, 11: 341-364.
- [17] Pelachaud C., M. Bilvi, Computational Model of Believable Conversational Agents, In *Communication in MAS: Background, Current Trends and Future*, Marc-Philippe Huget (Eds.), Springer-Verlag, 2003.
- [18] Russell J.A., a Circumplex Model of Affect, *Journal of Personality and Social Psychology*, 1980, 39(6) 1161-1178.
- [19] Strapparava C., A. Valitutti, WordNet-Affect: An Affective Extension of WordNet, in *Proc. of LREC'04*, Portugal, 2004.
- [20] Wojdel A., L.J.M. Rothkrantz, Parametric Generation of Facial Expressions Based on FACS, *Computer Graphics Forum*, Blackwell Publishing, 2005, 4:(24) 1-15.

ABOUT THE AUTHOR

Assoc.Prof. Dr. Drs. Leon J.M. Rothkrantz, Man-Machine Interaction Group, Delft University of Technology, Phone: +31 15 2787504, E-mail: l.j.m.Rothkrantz@ewi.tudelft.nl.
PhD Student Siska Fitrianie, MSc. PDEng, Man-Machine Interaction Group, Delft University of Technology, Phone: +31 15 2788543, E-mail: s.fitrianie@ewi.tudelft.nl.