Proceedings of the 3rd Workshop

Emotion and Computing – Current Research and Future Impact

Dirk Reichardt (Editor)

Kaiserslautern, Germany
September 23rd, 2008

ISSN 1865-6374

Emotion and Computing - current research and future impact

Workshop

The workshop focuses on the role of affect and emotion in computer systems including the three dimensions: emotion recognition, emotion generation and emotion modeling with special attention to AI specific problems and applications. Both shallow and deep models of emotion are in the focus of interest.

In recent years several approaches have been made concerning emotion recognition, emotion modeling, generation of emotional user interfaces and dialogue systems as well as intelligent and anthropomorphic communication agents.

There is a wide variety of motivations for emotional computing. From a scientific point of view, emotions play an essential role in decision making, as well as in perception and learning. Furthermore, emotions influence rational thinking and therefore should be part of rational agents as proposed by artificial intelligence research. Another focus is on human computer interfaces which include believable animations of interface agents.

One of the first interesting applications is a dialogue system which intends to generate natural human-like language and to react on emotional aspects of the utterances of the human partner adequately. Another field of application with significant AI influence is the field of computer games. An increasing interest in believable and intelligent computer characters can be recognized. As a third application type eLearning depends on the quality of learner and teacher interaction. Moreover, automatically recognizing the user's emotional state can help evaluating new products. In general many applications which integrate a user model often leave emotion and mood out of the adaptation loop. If we manage to cover this aspect, the quality of user adaptive systems can be increased significantly.

This workshop discusses the scientific methods considering their benefit for current and future applications. A special aspect of the discussion is the validation of emotionally biased systems and the emotion classification itself. The presented papers discuss theories, architectures and applications which are based upon emotional aspects of computing.

Organization and Scientific Committee

Organization

Prof. Dr. Dirk Reichardt, BA Stuttgart, University of Cooperative Education
http://wwwlehre.ba-stuttgart.de/~reichard

Scientific Committee

Dipl.-Ing. Joscha Bach, University of Osnabrück
http://www.cognitive-agents.org

Dr. Christian Becker-Asano, University of Bielefeld
http://www.techfak.uni-bielefeld.de/~cbecker/

Prof. Dr. Nicola Henze, University of Hannover
http://www.kbs.uni-hannover.de/~henze/

Dr. Manfred Kerber, University of Birmingham
http://www.cs.bham.ac.uk/~mmk/

Prof. Dr. Paul Levi, University of Stuttgart
http://www.ipvs.uni-stuttgart.de/ipvs/bv

Prof. Dr. J.-J.Ch. Meyer, University of Utrecht
http://www.cs.uu.nl/staff/jj.html

Dipl.-Ing. Christian Peter, Fraunhofer-Institut für Graphische Datenverarbeitung, Rostock
http://www.igd-r.fraunhofer.de/private/cpeter

Dr. Götz Renner, Daimler AG,
Head of Customer Research Center
http://www.daimler.com

Prof. Dr. Michael M. Richter, University of Calgary
http://people.freenet.de/mmrichter/

Dr. Marc Schröder, DFKI Saarbrücken
http://www.dfki.de/~schroed/

Workshop Program

The workshop provides talks covering a variety of aspects of emotion and computing to set a basis for a subsequent moderated discussion. The presentations are documented by reviewed short research papers.

Session 1: *9:00 – 10:00*

Welcome and Introduction

Identifying the common Ground of Emotion and Speech Acts
in View of Human-Robot interaction (HRI)
Wolfgang Gessner, Gesine Lenore Schiewer

Session 2: *10:30 – 12:00*

What is most important for an Emotion Markup Language?
Marc Schröder, Ian Wilson, William Jarrold, Dylan Evans,
Catherine Pelachaud, Enrico Zovato, and Kostas Karpouzis

Are You Happy with Your First Name?
Matthias Wimmer, Christoph Mayer, Martin Eggers, and Bernd Radig

Classification of facial Expressions based on Transient Features
Khadoudja Ghanem, Alice Caplier

Session 3: *13:30 – 15:00*

Demo *(all)*

About Emergence of Emotional Behaviour –
An Experiment using the Public Goods Game Scenario
Dirk Reichardt

Towards a platform for the education in emotion modeling based on virtual environments
Michal Bída, Cyril Brom

Session 4: *15:30 – 17:00 Discussion on Validation*

Correct Emotions? - how to work without ground truth

Moderated Discussion

Wrap up of Discussions and Workshop Summary

Scientific Research Papers

presented at the workshop

Identifying the common Ground of Emotion and Speech Acts in View of Human-Robot interaction (HRI)

Wolfgang Gessner & Gesine Lenore Schiewer

Abstract. Human emotions can serve as a model for user interfaces in the interaction of robots and humans (and vice versa). One of the dependent dimensions of emotions are speech act latencies, i.e. dispositions to react to specific situations by selected speech acts in emotional contexts. Emotions proper and speech acts are shown to be complementary types of reactions to cope with emotion-arising situations in a restriction model, which includes (according to the emotions given) some speech acts as consistent with these emotions, and others as inconsistent or at least implausible.

Keywords: Human emotions**,** Speech act theory, Situation representation, Androids, Human-Robot interfaces, Coping.

1. The common source of speech acts and emotions

Aiming at the integrative development of a structural theory of emotion and action resulting in a new communication medium at the user interface between android robots and humans, human emotion can serve as a model for analyzing, reconstructing and implementing the complex interrelations of mimics, gestures, verbal communication and elicitation of action schemes. Verbal communication as such is in our approach explicated as speech acts, which have been shown earlier to be rule-governed communicative activities depending on subjective representations of the objective situation given (Austin 1950, Searle 1969).

In some way comparable to this the elicitation of (cognitive) emotions can be understood as being triggered by (a) focal and topical cognitions of the subject and (b) by the subjective representation of situations already mentioned (Cf. Gessner 2004, ch. 7). So the utterance of speech acts (according to their constitutive rules) and the elicitation of emotions have a common base in relying on the same situation representation.

Furthermore, the constitutive rules of speech acts and the elicitation conditions of emotions can both be formalized in a 'language of mind' based on propositional attitudes (Cf. Gessner 2004 and Gessner & Schiewer's (2007) contribution to this workshop, '*Constructing Androids as Emotional Agents in Robot-Human Relationships'*). This amounts to the possibility of comparing the generative structure of emotions and of speech acts eventually corresponding to them on a common base, not only as regards to their contents respectively, but methodologically as well.

2. The logical form of speech acts in relation to emotions

Independent of where and whether a speech act will be executed, every speech act is disposed to a logical form which determines its success conditions and puts it into a relation to other speech acts. Searle and Vanderveken (1985, 2) aimed at presenting this form independently of the respectively different concrete possibilities of expression in the different languages. Based on this, a number of so-called illocutionary verbs of English were analysed in an abstract manner. In the context of illocutionary logic (which represents a further development of speech act theory) the following questions shall be answered (cf. Searle/Vanderveken 1985, 6):

- Which ones are the components of illocutionary force and which ones are the necessary and adequate conditions of the successful execution of elementary illocutionary acts? How can the success conditions of complex illocutionary acts be defined?

- What is the logical structure of the set of all illocutionary forces?

- Which logical relations exist between the different illocutionary types? Under which conditions does the execution of a speech act dispose for another illocutionary act?

The ‘constitutive rules’ of every speech act include the 'rules of propositional content', the 'preparatory condition’, the ‘sincerity condition' and the 'essential rule’. In turn all these rules can be described (or reformulated) in the language of mind LOM as well as the elicitation rules of emotions.

In detail, the rules of propositional content describe the type of content which is compatible with single speech acts, connecting the structure of our subjective world representations with the objective world: One can be annoyed about state of affairs and can be in a state of indignation by actions of other people, but can’t be in this emotional state concerning objective facts like the weather today (at least, if being minimally rational). Accordingly (as is well known from the classical texts on speech act theory) preparatory conditions define the conditions which must be given to make sense to utter a certain (type of) speech act at all. Furthermore, the essential rule of any speech act defines the communicative aim or target the speaker wants to achieve by uttering this single speech act.

It turns out that there are parallel relations between the propositional contents of speech acts and the intentional objects of emotions (i.e., the 'objects' as events, persons, conditions etc.) on which emotions focus. To give an example: ‘Admiration’ focuses on the person(s) which has been able to surprise us with actions not awaited by him or her (as the underlying subject of this emotion), ‘Indignation’ to the contrary. In general, it will depend on the ontology of world states, people’s actions and the interaction of both which should be, but has not yet been developed in a sufficient manner in contemporary psychology or cognitive science.

Furthermore, there are knowledge states, valuations, normative relations and ascriptions of ability which appear typically (but in changing combinations) in the elicitation formulas of every emotion, too [1]). This suggests the assumption that clues arise or have even coincidence rules derived from the parallelism of the mental conditions mentioned in constitutive rules of single speech acts and in the elicitations rules of single emotions, resulting in the 'fitting together' (compatibility) and also the incompatibility of single speech acts with special emotions. The minimal result of such an analysis are (as will be shown below) contingency tables between the 24 emotions analysed and the 41 speech acts analysed, as an amount of compatibility for the simultaneous appearance or as an amount of restriction for their incompatibility. Complex coincidence rules due to which strict connections between these two dimensions could be derived directly – based on comparisons between elements of the situation representation, the elicitation formulae for emotions and the constitutive rules for speech acts - would be the maximum result.

In his own attempt to answering these questions, Vanderveken (1990, 1) refers to the logically oriented philosophy of language in the tradition of Frege and Russell and emphasizes their contribution to the analysis of sentence meanings. Searle and Vanderveken (1985) have striven for a comparable formal development of speech act theory and have initiated a general logical theory of speech acts, later on called ‘illocutionary logic’. This logic contains recursive definitions of the set of all possible illocutionary forces as well as of the success and fulfilment conditions of elementary illocutionary acts. Vanderveken (1990, 1) applies the logical apparatus of his general semantics to English performatives and formulates translation rules for a great number of English performative verbs. The purpose of this translation consists in explaining the logical form of the illocutionary acts in question, the translation rules being based on the lexical analysis of these performative verbs.

For example, according to Vanderveken, the formalisation of the speech act ‚REQUEST ‘ (Aufforderung) is done in the following way:

> (16) *require*
> To require is to demand the hearer that he do something with the preparatory condition that it needs to be done.
> (Or, put otherwise, ‘to try to require something of somebody means to have reasons for demanding that’.)

Comparing Vanderveken’s (1991, 159) analysis with the formalisation of the speech act REQUEST (Aufforderung) using the ‘Language of Mind’ (LOM) and based on the earlier analysis of J. Searle will deliver the following structural results [2]):

---

[1] For details, see Gessner 2004 and the detailed analysis of elicitation conditions for emotions given there in ch. 7 and ch. 10.

[2] Signs and Symbols: $H =_{df.}$ Action, $C =_{df.}$ Ability, $B =_{df.}$ Belief, $K =_{df.}$ Knowledge, $I =_{df.}$ Intention, $W =_{df.}$ Wish, $\neg =_{df.}$ Negation, ©$\Rightarrow =_{df.}$ Cognitive implication, $G =_{df.}$ Present, $Z =_{df.}$ Future, $G \rightarrow Z =_{df.}$ Present to Future, $Sp =_{df.}$ Speaker, $Hr =_{df.}$ Hearer.

SPEECH ACT	REQUEST (AUFFORDERUNG)
Rules of propositional content	An action: $H_{Hr,Z}$
Preperatory conditions	(1) $C_{Hr,G\to Z}(H_{Hr,G\to Z})$ & $B_{Sp,G}(C_{Hr,G\to Z}(H_{Hr,G\to Z}))$ (2) $\neg K_{Sp,G}(I_{Hr,G\to Z}(H_{Hr,G\to Z}))$ & $\neg K_{Hr,G}(I_{Hr,G\to Z}(H_{Hr,G\to Z}))$
Sincerity condition	$W_{Sp,G}(H_{Hr,G\to Z})$
Essential rule	REQUEST ©$\Rightarrow$ $W_{Sp,G}(I_{Hr,G\to Z}(H_{Hr,G\to Z}))$

This formalisation corresponds to the original verbal formulation of the constitutive rules for ‚REQUEST' (modified, after Searle 1969). It turns out that the formalizations based on the formal instrument LOM are considerably more complex and will therefore allow for more meaningful results. Furthermore we rely on the careful and detailed semantic speech act analyses in the meaning analysis of selected performative verbs by Wagner (2001) which he gave by using all the relevant and appropriate research literature in this topic for German. They are the basis of the explication of the constitutive rules of the speech acts taken into account as well as of the formalizations carried out on this basis.

3. Interrelations and restrictions between emotions and speech acts

In this chapter we will show the complex interrelations of uttering speech acts with having emotions as classes of mutual implication, of compatibility and of restriction as being based on functional considerations.

At first it has to be adhered that the marking of the illocutionary force of a speech act indexes the propositional involvement of the speaker to the object expressed in the proposition. The speaker is sad and not happy with the existence of the condition in question which is expressed in this proposition (cf. Vanderveken 1990, 15). For example, somebody is 'complaint' or disposed to 'lamentation' if he is at the same time sad about an existing fact which is in parallel expressed in the propositional content of this speech act. By this way the specific components of the illocutionary force can be inferred back on the emotional assessments. Conversely, little modifications of this proposition may alter even the explicit illocutionary force of the whole sentence, including the explicit meaning of the speech act's illocutionary verb. This is due to the possibility of building 'indirect' speech acts, a matter which can't be dealt with here for reasons of space.

Nonetheless, speech act theory delivers a methodical means to contribute to the investigation of the relations of emotions and speech act choice:

- This will result primarily from the analysis of the sincerity conditions of speech acts: Any successful execution of an illocutionary act constitutes the expression of a mental state with a psychological mode referring to a state, which is described by the propositional content of this speech act and thereby is determined as its intentional object. For example, a speaker giving a promise expresses his intention to execute an act which corresponds to what he promises (Principle of parallelism of mental attitudes).

- Secondly, our mental states are always (at least in principle) expressible by illocutionary acts. For example we can express beliefs in statements or assertions and wishes in inquiries, interpellations or requests. (Principle of illocutionary expressibility).

- Finally, the direction of fit of the mental state expressed is in general identical with the one given in the illocutionary act which is executed in order to express this mental state. Furthermore, the conditions of fulfilment of the illocutionary act and the mental act are logically connected (Principle of identity of fit).

In any case the mental state of the emitter will play a crucial role for the execution of the corresponding illocutionary act (or acts). So it seems plausible to impute a normative connection and an empirically ascertainable covariance between these two dimensions of single emotions and special speech acts. The connection to be captured normatively is based on the existence of common elements of the situation representation in (a) the elicitation rules of emotions and (b) the constitutive rules defining the components of the illocutionary force in question, respectively.

For example a negative situation assessment, that is the injury of justifiable expectations and the damage done to the own person, can lead to the elicitation of the emotion ANNOYANCE / IRRITATION, the german 'Ärger' (cf. Gessner 2004, 128 ff.). The components of illocutionary acts possibly used here like 'complaining' correspondingly imply the expression of the displeasure in the area of the sincerity condition and in the area of the preparatory condition that the facts in question - the object of this trouble - are judged negatively. In the case of the illocutionary act 'moan' the aspect of deep sadness additionally enters in the sincerity condition there (cf. Vanderveken 1990, 181).

By this, an immediate coupling arises between the selected subset of robot-relevant emotions [3]) and the choice of illocutionary acts. Furthermore it can be expected that –

---

[3] 24 emotions have been selected, which (in our view) are necessary and sufficient to represent the 'mental space' and the corresponding communicative and action latencies broad enough to cover the 'needs' and intentions connected with HRI-contexts. The (rather complicated) functional reasons for this selection can't be explained here.

according to a sort of reverse conclusion – some specific emotions are imcompatible with some specific speech acts.

So it would not be plausible that in the case of the emotion ANNOYANCE / IRRITATION (which shows the negative assessment of the actions of another person in the preparatory condition) that this person is *praised* under preservation of the sincerity condition, because 'praise' expresses a high degree of approval (cf. Vanderveken 1990, 215) and is thus incompatible with this emotion regarding the same action of this other person.

These interrelations can be represented schematically in individual analyses of the relations and restrictions between 24 emotions and 41 speech acts (cf. Schiewer 2007), demonstrated here by the example of the relations and restrictions of the performative verb or speech act type REQUEST (Aufforderung):

REPUGNANCE (ABSCHEU)	Request to Hr to cancel a future action, if it corresponds to the intentional object of the emotion
ANNOYANCE/IRRITATION (ÄRGER)	Request to Hr to cancel a future action, if it corresponds to the intentional object of the emotion
SUSPICION (ARGWOHN)	None
REGRET (BEDAUERN)	None
ADMIRATION (BEWUNDERUNG)	Request to Hr to continue with the intentional object of the emotion and not to cancel it
GRATITUDE (DANKBARKEIT)	Request to Hr to continue with the intentional object of the emotion and not to cancel it
INDIGNATION (EMPÖRUNG)	Request to Hr to cancel a future action, if it corresponds to the intentional object of the emotion
HORROR (ENTSETZEN)	Request to Hr to cancel a future action, if it corresponds to the intentional object of the emotion
DISAPPOINTMENT (ENTTÄUSCHUNG)	Request to Hr to execute an action if Sp believes Hr is able to this in spite of the fact Hr had acted up to now to the contrary
RELIEF (ERLEICHTERUNG)	None
JOY (FREUDE)	None
FEAR (FURCHT)	Request to Hr (or third person) to cancel their planned or ongoing action
HAPPINESS (GLÜCK)	None
HOPE (HOFFNUNG)	Request to Hr to allow for he result hoped for or to bring It forward by an own action
BOREDOM (LANGEWEILE)	Request to Hr to alrer an existing state if Sp believes Hr Is responsible fort his state
COMPASSION (MITLEID)	Request to a third person in order to bring it about that an action is done (or omitted) if Sp believes that this action could be harmful to Hr (or he believes that the forbearance of this action could prevent this harm from Hr)
SHAME (SCHAM)	None
FEELING OF GUILT (SCHULDGEFÜHL)	None
SORROW (SORGE)	Request to Hr to cancel a future action of him if Sp believes H is able to execute this action

PRIDE (STOLZ)	None
SURPRISE (ÜBERRASCHUNG)	None
ANTICIPATION (VORFREUDE)	None
ANGER (ZORN)	Request to Hr to fulfill an action which compensates the harm done to Sp by Hr, or request to Hr not to repeat this harming action
SATISFACTION (ZUFRIEDENHEIT)	None

4. Giving an example: The relations of speech acts and emotions in ‚annoyance / irritation'

Dispositions for speech acts result on the basis of subjective situation representations. The 'triggering' of speech acts can be understood (quite similar to the elicitation of emotions) as a process of sample comparison between the given set of elements of this situation and the 'constitutive rules' for speech acts as a total set of predefined cognitive patterns. The elicitation formulae for individual emotions (cf. Gessner 2004, Ch. 7 and Ch. 11) as well as the constitutive rules of specific speech act types are to be 'tested' by subjects with respect to their 'fit' with the respective situation description. This comparison process leads the 'triggering' of an emotion and one (or more) speech act types (as a language generation process) by filling these abstract cognitive schemes as 'open variables' respectively with situation-specific contents and elements. So speech acts are (as explicitly communicative reactions on 'problematic' situations) in parallel with emotions in coping with this 'problematic' situation. Just as forecasting the appearance of the emotion altogether is possible, the validity of the respective samples of speech acts can be forecasted likewise under the prerequisite of a situation representation given subjectively which fits the triggering schemata of emotions and speech acts as well.

In the context of the given example of the emotion ‚ANNOYANCE / IRRITATION' essentially the following speech acts REQUEST (Aufforderung), WARNING (Warnung), ADMONITION (Mahnung), COMPLAINT (Beschwerde), THREAT (Drohung) should be examined according to the criteria given above for their compatibility and integration with this emotion. In the case of the speech act ‚REQUEST (Aufforderung)' (compare its constitutive rules given above!) the following comments can be given concerning its compatibility with the elicitation conditions of the emotion 'annoyance/irritation':

The propositional content of the speech act 'REQUEST' and the intentional object of the emotion 'ANNOYANCE / IRRITATION' are identical, namely the problematic action of the other person. The negative evaluation of this action in the emotional context derives from the failed obligation to the execution of this action in the context of 'ANNOYANCE / IRRITATION'. In the speech act context this action is (in the introduction rule and in the sincerity rule) presupposed in the default case as an action assessed positively by the potential speaker. The constitutive rule should be extended by 'REQUESTING' for cases of the failure of that action. Similarly for the essential rule of this speech act, to the effect that the potential speaker wants the hearer to form an intention to forbear an action not requested by the potential speaker.

In this respect the speech act 'REQUESTING' is primarily compatible with this emotion in those cases in which the intention of the other person to perform an action undesired by the speaker exists and is known to exist, but has not yet been executed, still lies ahead or has just begun being executed.

Using this case as an analytical example, the other speech act types just mentioned can be analysed in a comparable manner, using variants in speakers' and hearers' intentions and valuations, referring to variants in time relations of actions and their consequences and to the state of the consequences as still possible or already stated, i.e. still preventable or already irreversible. The result of these more fine grained analysis will deliver a matrix of compatibility and incompatibility of any singular emotion with the considered speech acts, and by this the relative impacts and the relations of emotion and speech in coping with situations can be determined.

So we have shown that speech acts (as utterances adapted to situations) and emotions as ‚quick and dirty' reactions to situations fulfil parallel and mutually complementary functions, which allow for people – and later on android robots, too – to cope in a systematic and successful way with the ever changing situations they meet in variable circumstances.

Bibliography

Austin, J. L. (1950). How to do Things with Words. Oxford: Clarendon Press.

Gessner, W. (2004). Die kognitive Emergenz von Emotionen. Paderborn: Mentis.

Rolf, E. (1997). Illokutionäre Kräfte. Grundbegriffe der Illokutionslogik. Opladen: Westdeutscher Verlag.

Gessner, W. & G. L. Schiewer (2007). KI 2007, Workshop ‚Emotion and Computing': Constructing Androids as emotional Agents in Human-Robot Relationships.

Schiewer, G. L. (2007). Project MMI 1978, Working Paper 25: Kontingenztabellen möglicher Relationen und Restriktionen zwischen 41 Sprechhandlungen und 24 Emotionen

Searle, J. (1969). Speech Acts. An essay in the Philosophy of Language.

Searle, J. R. & D. Vanderveken (1985). Foundations of Illocutionary Logic. Cambridge: Cambridge University Press.

Vanderveken, D. (1990/1991). Meaning and Speech Acts. 2 Vols. Cambridge: Cambridge University Press.

Wagner, K. R. (2001). Pragmatik der deutschen Sprache. Frankfurt am Main et al: Peter Lang.

What is most important for an Emotion Markup Language?

Marc Schröder[1], Ian Wilson[2], William Jarrold[3], Dylan Evans[4], Catherine Pelachaud[5], Enrico Zovato[6], and Kostas Karpouzis[7]

[1] DFKI GmbH, Saarbrücken, Germany
[2] Emotion AI, Cambridge, UK
[3] SRI International, Menlo Park, CA, USA
[4] Cork Constraint Computation Centre, University College Cork, Ireland
[5] University of Paris 8, INRIA, France
[6] Loquendo, S.p.A., Turin, Italy
[7] Image, Video and Multimedia Systems Lab, Nat. Tech. Univ. Athens, Greece
`http://www.w3.org/2005/Incubator/emotion`

Abstract. The paper describes the consolidated set of mandatory and optional requirements for an Emotion Markup Language as identified by the W3C Emotion Markup Language Incubator group. It also exposes a number of relevant questions that arise from theoretical considerations as well as from the intended use of the markup language.

1 Introduction

Despite the growing relevance of affective computing in technological environments, there is still no common interchange language for applications that deal with emotion and affect. The W3C Emotion Incubator Group was established at the beginning of 2006 with the aim to fill this gap. The group, composed of W3C members and of invited experts from industry and academic institutions, aims to design a general-purpose emotion markup language that can be effectively used in different contexts and applications. Due to the large variety of emotion representations used both in scientific theory and in applications, the language must be compatible with a broad range of emotion theories, and must be customisable to some extent within a standard framework. Furthermore, the language must interoperate with other markups, particularly with those that are standards in multimedia applications.

The first year of the group's work was focused on defining the requirements of the new language [1, 2] on the basis of a bottom-up collection of use cases. Current work is moving towards a concrete specification for the language. Given the long initial list of requirements, it was necessary to distinguish between mandatory ("must have") and optional ("should have") requirements. 14 experts from industry and academia evaluated each of the 22 initial requirements according to its importance [3]. Agreement was high in most cases, leading to a natural separation into two groups. Borderline cases, including *action tendencies* and *regulation*, were discussed in the group before making a choice.

The present paper presents the outcome of the prioritisation on requirements, and reports on discussions of both theoretical and usage considerations.

2 "Must have"s for an EmotionML

This section describes the markup elements that are considered mandatory by the Emotion Markup Language Incubator Group. They comprise a sophisticated description of the emotion or related phenomenon itself; linking mechanisms to other items in the world; meta-information about the emotion annotation; and a generic notion of global metadata.

2.1 Emotion Core

The description of the emotion or related state itself naturally receives the most prominent place in an emotion markup language. Most of the items on the original requirements list [2] are considered mandatory, allowing the user to represent most of the multiple facets of an emotion [4], but also less intense affective states [5].

- **Core 1. Type of emotion-related phenomenon**

First, the language must allow the user to be explicit about the type of phenomenon that is being represented: an *emotion* in the strong sense, i.e. a momentary, intense episode triggered by a concrete event, or rather a mood, an attitude, an interpersonal stance, etc.

Which taxonomy to use for distinguishing types of emotion-related phenomena is an open research question, and as for other elements of the emotion markup language, any standard will only be able to propose a "default" answer. Users must be able to replace the suggested taxonomy with one that fits their own needs. One possible starting point for proposing a default set, from the literature on emotion theory, is provided by Scherer [5].

- **Core 2. Emotion categories**
- **Core 3. Emotion dimensions**
- **Core 4. Appraisals related to the emotion**
- **Core 5. Action tendencies**

Emotions can be described in a number of different ways, using categories, dimensions, appraisals, or action tendencies. Details of these have been described elsewhere (e.g., [4, 2]), so we only point out some issues relevant for defining a markup language.

Emotion categories can be chosen from a set of discrete labels. Dimensions, appraisals and action tendencies seem to be best represented as "scale" values. This raises interesting questions, see Section 4.1.

Each of the four types of representation needs a vocabulary of names for the categories, dimensions etc.; again, the aim will be to propose a meaningful "default", and allow users to use a different set if they have specific needs.

- **Core 6. Multiple and/or complex emotions**

Multiple emotions may be co-occurring in the same experiencer, e.g. when a person is angry about one thing and sad about another, when the face shows one emotion and the voice another, or in cases of regulation (see Core 9, below).

– **Core 7. Emotion intensity**

The intensity of an emotion is a unipolar scale; the question how to represent scales is not easily answered, though, see Section 4.1 for a short discussion.

– **Core 8. Emotion timing**

The temporal scope of an emotion markup may be defined through a combination of start and end times, or by linking to items located on the time line such as utterances or gestures. The *time course* of an emotion markup may be defined through a sampling mechanism, providing values at fixed intervals.

2.2 Meta-information about emotion annotation

– **Meta 1. Confidence / probability**

Both machine classifiers and human annotators need to indicate the degree of confidence that a certain element of the representation is correct.

– **Meta 2. Modality**

Emotion may be expressed specifically in a certain modality, e.g. face, voice body posture or hand gestures, but also lighting, font shape, etc.

2.3 Links to the "rest of the world"

Emotion markup is always about something. Providing suitable links to external entities is essential for the interpretation of the emotion markup.

– **Links 1. Links to media**
– **Links 2. Position on a time line in externally linked objects**

A generic linking mechanism is envisaged. A link may point to a media object, such as a picture, an audio or video file, or a node in an XML document; this may be complemented with timing information, such as a start time and a duration.

– **Links 3. The semantics of links to the "rest of the world"**

Links must be assigned a meaning. Initially, the following meanings are envisaged: the experiencer (who "has" the emotion); the observable behaviour "expressing" the emotion; the trigger, cause or eliciting event of the emotion; and the object or target of the emotion (i.e., the thing that the emotion is about).

2.4 Global metadata

In order to facilitate communication between a producer and a consumer of emotional data with respect to application-specific information, the emotional markup may need to contain global metadata.

– **Global 0. A generic mechanism to represent global metadata**

In the Must Have section, we have previewed only a relatively unspecific placeholder (hence the identifier "Global 0") for the various more specific but currently optional requirements for global metadata (see below).

3 “Should have”s for an EmotionML

The elements in this section are considered less important and urgent; if their implementation poses non-trivial problems, a first draft of the emotion markup language does not need to implement them. Nevertheless, they are needed for certain use cases, and should be added in future versions of the markup language.

3.1 Emotion Core

– **Core 9. Emotion regulation**

Regulation covers a range of manipulations of an emotion or its expression by the experiencer. In a very basic interpretation, this includes a difference between the internal and the externalised state, i.e. cases of simulation and suppression. However, considerably more complex models of emotion regulation are described in the literature (e.g., [6]), and a suitable degree of abstraction will need to be found; see also the discussion in Section 4.1.

3.2 Meta-information about emotion annotation

– **Meta 3. Acting**

Specific annotations are needed for describing the properties of acted material, such as perceived naturalness, authenticity, quality of acting, etc.

3.3 Global Metadata

– **Global 1. Info on Person(s)**
– **Global 2. Social and communicative environment**
– **Global 3. Purpose of classification**
– **Global 4. Technical environment**

Several specific kinds of global metadata are needed in different use cases in order to properly interpret the emotion markup in context. For some kinds of metadata (e.g., information about persons), it may be possible to reuse existing metadata annotation schemes; others (e.g., social environment) are specifically relevant in the context of emotion markup, and will need to be modelled explicitly. Purpose of classification and technical environment are needed for the emotion recognition use case, where they are needed to interpret the markup.

3.4 Ontologies of emotion descriptions

– **Onto 1. Mappings between different emotion representations**
– **Onto 2. Relationships between concepts in an emotion description**

Different emotion representations, and different concepts within an emotion representation, are not independent of one another; if their relation could be made explicit, that would allow for mappings (such as, locating an emotion category on dimensional scales), or for better interpretations (e.g., by making the similarity or difference between emotion categories explicit).

It remains to be seen whether such relationships are best modelled within the emotion markup language, or whether it is better to model them as a complement to the markup language.

4 Discussion

The targeted specification raises a considerable number of issues, combining questions of theoretical interest with practical considerations. The present section outlines a number of key questions that have been raised in the group.

4.1 Theoretical Issues

The Status of Action Tendencies Action tendencies are listed among the possible descriptions of emotions that are mandatory for the EmotionML. The concept stems from emotion theory – for example, according to Frijda [7, p. 88, Table 2.1], desire is linked to a tendency to approach, fear is linked to a tendency to avoid, and so on. Action tendencies are potentially very relevant for use cases where emotions play a role in driving behaviour, such as in the behaviour planning component of non-player characters in games, or in robot companions.

However, there are some interesting difficulties with this requirement. One of them is the dependency of action tendencies on the effectors available to the system, which have huge variation. As an example, an action tendency of "eat food" would make sense for systems capable of (simulating) eating whereas a more informational "consume energy" might be a more useful and wide ranging description that would equally apply to robots that recharge their batteries.

Another difficulty concerns the distinction between action tendencies and other forms of observable emotion behaviour such as facial expression. It may seem arbitrary to include a specific requirement for action tendencies in the markup language, while all other information about the observable behaviour "expressing" the emotion is relegated to a subcategory of metadata that specifies the semantics of links to the "rest of the world". However, there are sound practical reasons for this decision. For one thing, there are well-established standards for certain visual expressions of emotion that have already been specified outside of the proposed markup language, such as MPEG-4, H-anim, and FACS, while no such 'standard' representation language currently exists for behaviours. Also, there are indefinitely many forms of emotional expression, from facial expressions, physiological parameters in humans and humanoid agents to colours and flashing lights in robots. This is generically unlimited, and so it was decided to leave this information outside the emotion markup language, in order to keep the language reasonably compact.

Complex Emotions and Regulation Complex emotions are listed as a mandatory requirement for the EmotionML. This enables the representation of cases where an event may be evaluated from different perspectives leading to the superposition of two emotions, as well as cases where different emotions are apparent in different modalities.

Regulation, however, is listed as an optional requirement, despite the fact that it is an important source for complex emotions. For example, the display of an emotional state may be impeded due to some socio-cultural rules [8]: the expression of one emotion may be masked by another one, it may be inhibited, minimised or even exaggerated. Alternatively, it is possible to regulate, to some extent, the emotion itself rather than its expression, through a process of re-appraisal [6]. Representing regulation in a scientifically appropriate way appears to be a non-trivial challenge; for this reason, we avoided making regulation a mandatory requirement for EmotionML, despite its importance for modelling certain types of complex emotions.

Scale Values Several of the elements listed in Sections 2 and 3 seem to be best represented as values along some sort of scale. A straightforward representation of scale values, and an obvious candidate for a "default" representation, would be a continuous unitless scale such as [0,1]. Such a standard value range would have the advantage of easily supporting interoperability between technological components; however, it would limit the use of EmotionML in several ways.

One potential deficiency of [0,1] is that it may be overly restrictive for representing emotion intensity. For example, is there some maximum amount of experienceable joy or despair? Independently of that theoretical question, it may be desirable for some applications to generate exaggerated values for emotions and their expressions, e.g. for cartoon animation.

Furthermore, human annotators often use a set of discrete labels such as a five-point Likert scale ("strongly disagree", "disagree", ..., "strongly agree"). Such scales are at best ordinal, but it would be misleading to map them to a metric interval such as [0,1] – the psychometric difference between "neither agree nor disagree" and "agree" cannot be assumed to be as large as the difference between "agree" and "strongly agree". Any such mapping would necessarily imply a loss of information and of accuracy.

Finally, human judgements may be class-specific (e.g., "For a child, Young Johnny exhibited a low tendency to avoid strangers."), and may provide only partial ordering (human raters may state consistently that $A > B$ and $B > C$ but may not agree that $A > C$).

These considerations show that a seemingly simple aspect of the language, such as the values on a scale, raise complex questions in view of both scientific validity and general usability.

4.2 Usage issues

Target Audience The participants in the group represent a broad cross-section of users, and as such are in themselves a reasonable reflection of the potential user base for an EmotionML. These user groups come under two general categories, industry and academia, and their potential products, services and research can broadly be defined as emotion annotation, emotion recognition and emotion generation, as we have previously defined in our representative use cases [1]. We can further summarise the use cases into two groups, data storage (annotation) and data transmission (recognition and generation).

This is a wide spectrum of potential uses and it is a challenge for the group to define a standard that is flexible enough to be suitable for the requirements of each user group and functional task. While academia and industry can and should share standards, users who require a data storage standard and those who require a data transmission standard may have quite different requirements. Furthermore, any emerging standard should respect and support existing processes concerning data extraction, storage and manipulation; if it does not, then commercial adoption will be minimal and as a result the standard may be considered irrelevant.

Future Proofing Working from a well defined set of use cases has allowed us to better understand the requirements of potential users and form constraints that define how far we expect the standard to stretch. However, the standard should also be flexible enough to accommodate future uses that we have not currently envisioned.

To this end, one of our guiding principles has been to define only a core standard but allow and encourage the core to be extended by the use of custom vocabularies, notably for the emotion description itself. We aim to produce, within the core specification, a set of default values for categories, dimensions, appraisals, etc., appropriate to the transmission and storage of emotion data that will serve many, perhaps non specialist, users needs. These can, however, be replaced or "overloaded" by the user's own defined set.

This extensibility should give the standard a longer life span and avoid rapid obsolescence.

Interoperability Another centrally important objective of the standard is to enable or facilitate interoperability between systems that process and/or store emotional data. In this respect, EmotionML faces a number of challenges. Its main challenge is that within the area of emotion research there are currently no agreed and accepted standards for concepts such as categories, labels, dimensions and scales. To enable interoperability it will be necessary for us to define a mechanism that can either translate between different formats or enforce a set of core data types within the standard and then allow users to translate from this to their own preferred formats. In this way any transmission of data is guaranteed to be in a standard format and so the recipient need not worry or know about the sender's internally used format.

4.3 Towards a customisable standard?

The above discussion has highlighted the diversity of potential uses of the Emotion Markup Language in different contexts. To accomodate these, the language will have to provide sufficient flexibility: to describe emotions using one or several descriptors (categories, dimensions, appraisals and/or action tendencies); to use a pre-defined or a custom vocabulary for a descriptor (e.g., a custom list of emotion words or of appraisal dimensions); discussion is ongoing whether to provide users with the choice among one of several ways to indicate scale values – continuous bounded or unbounded, pre-defined or custom ordered lists, etc. At the same time, the discussion has pointed out the need to use standard representations for exchanging data between technological components.

The task of the group is to weigh the need for standardisation, required for interoperability, against the need for customisation, required for representing the concepts with which potential users are already working nowadays. The guiding principle used is to provide a choice only where it is truly required; to propose default sets for every choice; and to introduce a mechanism for mapping between representations where possible.

5 Conclusion

The paper has presented the mandatory and the optional elements of an emotion markup language, and has discussed a number of key aspects to take into account when designing the language.

The next step will now be the formulation of a concrete specification draft, and to collect feedback from potential users, in order to ensure that the language can actually be used as intended: as the most carefully designed and most generally usable emotion markup language to date.

Acknowledgements

Preparation of this paper was supported by the W3C and has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 211486 (SEMAINE).

References

1. Schröder, M., Devillers, L., Karpouzis, K., Martin, J.C., Pelachaud, C., Peter, C., Pirker, H., Schuller, B., Tao, J., Wilson, I.: What should a generic emotion markup language be able to represent? In: Proc. 2nd International Conference on Affective Computing and Intelligent Interaction (ACII'2007), Lisbon, Portugal (2007)
2. Schröder, M., Zovato, E., Pirker, H., Peter, C., Burkhardt, F.: W3C Emotion Incubator Group final report (2007)
3. Burkhardt, F., Schröder, M.: Emotion markup language: Requirements with priorities. W3C Incubator Group report (2008)
4. Cowie, R., Cornelius, R.R.: Describing the emotional states that are expressed in speech. Speech Communication Special Issue on Speech and Emotion **40** (2003) 5–32
5. Scherer, K.R.: Psychological models of emotion. In Borod, J.C., ed.: The Neuropsychology of Emotion. Oxford University Press, New York (2000) 137–162
6. Gross, J.J.: Emotion regulation in adulthood: Timing is everything. Current Directions in Psychological Science **10** (2001) 214–219
7. Frijda, N.H.: The Emotions. Cambridge University Press, Cambridge, UK (1986)
8. Ekman, P.: Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life. Times Books (2003)

Are You Happy with Your First Name?

Matthias Wimmer, Christoph Mayer, Martin Eggers, and Bernd Radig

Institut für Informatik, Technische Universität München, Germany

Abstract. Psychological and sociological insight is often based on manually evaluating data, which this is a laborious and time-consuming task. Inspecting data, such as images, by computer systems provides the opportunity to consider much more information. This paper presents a system, that is able to automatically analyze images with human faces robustly and accurately and infer useful information.
As an example, we conducted a large-scale survey on a huge data set of images collected from the Internet. This experiment aims at determining a relationship between people's first names and their magnitude of smile visible in the images.

1 Introduction and Motivation

Folk psychology assures us that many aspects of first impressions are vitally important, such as clothes or hairstyle [10]. One characteristic we initially present is our first name. Parents usually select their children's first names with great care, which shows that this characteristic is considered to be very important and partly represents our personality. Psychologists state that there is evidence that first names influence the character building of human beings. Sociologists conduct research upon the impact of our first names on our behavior and our status in society [5, 9, 21]. Scientific insight into this area is usually obtained by conducting surveys questioning people about their opinion. Lawson [12] obtained a ranking of first names w.r.t. liked ones and disliked ones. Fryer et al. [6] investigate the social interaction of distinctively black names. Steele et al. [10] asked 80 volunteers to describe 8 first names by a set of 5 bipolar adjectives. Their results identify a relationship between first names and first impressions.

Problem Statement: Most psychological and sociological research is currently based on subjectively evaluating images and other kinds of data. However, human ability to inspect images manually is limited to a certain number only. This restriction prevents the drawing of objective conclusions. In contrast, inspecting images by computer systems provides the opportunity to consider a much larger amount of data. However, the restricted accuracy of computer algorithms made their application impossible in the past.

Solution Outline: Due to recent advances in the field of Computer Vision and the field of Machine Learning, we are now able to automatically analyze images depicting human faces both more robustly and more accurately. This paper describes the components of a computer system that automatically inspects a

Fig. 1. Our system automatically determines the facial configuration and infers the magnitude of the smile visible.

large number of images with human faces and extracts useful information about displayed emotion, such as the magnitude of the smile visible.

Contributions: The research field on investigating social behavior of first names is now given the opportunity to automatically analyze a large number of images. The large amount of data renders its results more reliable and the utilization of Machine Learning provides more objectivity. We provide the Computer Vision community with a large data base of frontal faces that are partly annotated with ground truth. This data base can be used for training, evaluation, and benchmarking purposes. Similar psychological and sociological experiments that require the analysis single images can be conducted in a very similar manner.

In order to demonstrate the benefit of our system, we conduct a well-known sociological experiment, an automatic evaluation of first names related to the degree of happiness visible in facial expression, see Figure 1. For this reason, we download a very large number of images from the web that are tied to different first names. Afterwards, we automatically analyze the magnitude of the smiles within these images.

This paper continues as follows: Section 2 roughly describes recent technical achievements in the domain of model-based image interpretation, which are essential for the success of our system. Section 3 explains how we derive the magnitude of happiness from visual information in detail. Section 4 describes the experiment on analyzing the smile visible within a large number of images.

2 Face Model Fitting

Model-based techniques allow to accurately interpret images with faces by exploiting a priori knowledge of human faces, such as shape and texture. They reduce the large amount of image data to a small set of model parameters, which facilitates and accelerates the subsequent interpretation process. Fitting the face model consists of the computational challenge of finding the model parameters that best describe the facial configuration within a given image. This section presents the four main components of model-based techniques, as described in [18].

The face model reflects the configuration of the visible face using a vector of parameters $\mathbf{p}$. We integrate a deformable Active Appearance Model (AAM) [3]. Its parameters $\mathbf{p} = (t_x, t_y, s, \theta, \mathbf{b}, \mathbf{u})^T$ describe the affine transformation, the

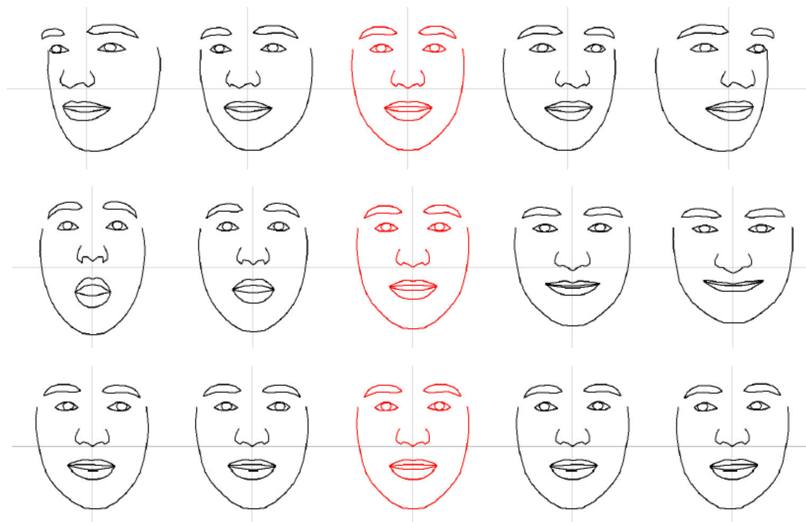


Fig. 2. Changing the deformation parameters **b** yields highly semantic constitutions of the face. $\mathbf{b}_1$ rotates the head. $\mathbf{b}_3$ opens the mouth. Lower-most row: $\mathbf{b}_9$ moves pupils accordingly.

deformation **b**, and the texture **u**. Figure 2 shows how **b** relates to the deformation of the face. In contrast, **u** returns evidence of the facial appearance, such as complexion, facial hair, wrinkles, etc.

The localization algorithm computes an initial estimate of the model parameters that is further refined by the subsequent fitting algorithm. Our system integrates the approach of [17], which detects the model's affine transformation.

The objective function yields a comparable value that specifies how accurately a parameterized model matches an image. Traditional approaches manually specify the objective function in a laborious and error-prone task. In contrast, we use our method introduces in [19], which automatically learns the objective function from a large set of training data based on objective information theoretic measures. This approach yields more robust and accurate objective functions, which greatly facilitate the task of the associated fitting algorithms. In turn, precisely fitted models ensure the high accuracy of facial expression recognition.

The fitting algorithm searches for the model parameters that best describe the face visible in the image. They correspond to the minimum of the objective function. Fitting algorithms have been the subject of intensive research and evaluation, see [8] for a recent overview. We adapt the objective function rather than the fitting algorithm to the specifics of the face interpretation scenario, and therefore, we are able to use any fitting algorithm. Since emotion interpretation mostly requires real-time capabilities, our approach integrates a quick *hill climbing* algorithm. Note that the reasonable specification of the objective function makes this local optimization method nearly as accurate as global optimization strategies, such as a *Genetic Algorithm*.

3 Computing the Magnitude of a Smile

According to the state-of-the-art, we do not manually specify the classifier that computes the magnitude of the smile, because this would be a laborious and error-prone task. Instead, we use a Machine Learning technique, which learns the calculation rules from a large number of training images that are annotated with the intended magnitudes of the smile m. We integrate tree-based regression (M5') [14, 20]. Note that this computation is conducted for single images and not for sequences.

3.1 Generating Training Data

Our training data set is the Cohn-Kanade data base [16], because it represents a general benchmark dataset in the area of facial expression recognition. It contains 488 short image sequences consisting of between 4 and 66 images, with each of the sequences displaying one of the six universal facial expressions [11]. Each sequence starts with a neutral face and develops into the apex expression. We fit the face model to each of the 6110 images and compute several features for each image individually.

Two aspects turn out to be most descriptive for facial expressions: the shape of the face and its texture. The face model described in Section 2 has been chosen such that it contains descriptors for both aspects, namely the deformation parameters $\mathbf{b}$ and the texture parameters $\mathbf{u}$. They represent the features from which we derive the magnitude of the smile.

Furthermore, Machine Learning techniques require to provide a correct result value which indicates the magnitude of the smile in our case. In order not to entirely specify this value by hand, we conduct a partly automated process, which only requires us to specify the facial expression visible in the entire sequence. The magnitude of the smile m as it is visible on a specific image is computed according to Equation 1 by linear interpolation between the first frame and the last frame of the image sequence.

$$m = \begin{cases} \frac{\text{index of current frame}}{\text{length of sequence}} & \text{sequence shows a smiling face} \\ 0 & \text{sequence shows other facial expression} \end{cases} \tag{1}$$

3.2 Precision of the Classifier

The precision of computing the magnitude of happiness influences the experiments performed. Therefore, we evaluate its accuracy by conducting a 10-fold

classifier	happiness	sadness	surprise	anger	fear	disgust
2-class classification (distinguishes between neutral and expression)						
J48	96.23%	90.51%	94.37%	87.50%	92.76%	87.78%
SVM	97.57%	96.27%	98.87%	92.61%	96.90%	95.00%
Neural Network	97.84%	97.29%	98.31%	92.05%	97.59%	94.44%
regression (computes the magnitude of an expression)						
M5'	35.00%	45.77%	31.43%	48.29%	59.56%	52.66%

Table 1. This table illustrates the accuracy of the classifiers and the regressors tested. Note that the classifiers are judged by the correctly classified instances, whereas the regressors' accuracies is measured by the relative absolute error.

cross-validation. This kind of analysis splits the training database into 10 equally large parts. During 10 iterations, 9 parts are used for training the algorithm and the remaining part is used for testing its accuracy. The overall accuracy represents the mean of the 10 iterations. Table 1 illustrates the accuracy of different techniques for classification and regression.

Note the difference to related techniques that benefit from analyzing an entire image sequence [1, 13, 15]. In contrast, we use the data of single images only. Other approaches analyze the neutral and the apex expression only, whereas we compute a percentage value that gives evidence about the magnitude of the facial expression. All these differences imply the area of application and different degrees of accuracy.

4 Experimenting with First Names

This section describes an example experiment that requires to inspect a large number of images. It investigates the relation between the magnitude of the visible smiles and the first names of the persons depicted. Conducting this experiment automatically allows to analyze much more images while saving a lot of man power and therefore provides more consolidated findings. This experiment requires to gather a list of first names and a lot of images with faces for each of these first names. Afterwards, our automatic system analyzes each image and computes the magnitude of the smile.

We obtain 500 German first names (gender-balanced) from the web [2]. For each of these first names, we query the Google image search [7] for face images with this name. We collect around 20 high resolution images for each first name. Note, that the Google image search is queried to return images that contain faces only.

We calculate statistical measurements such as mean and deviation from the classification results for every first name. Manually comparing these statistical measurements for 500 first names shows that the classification results correspond to human intuitive estimation. Figure 3 presents the distribution of the magnitude of the smiles across the different first names. Our image base contains approximately 5000 male and 5000 female faces. Comparison of the mean

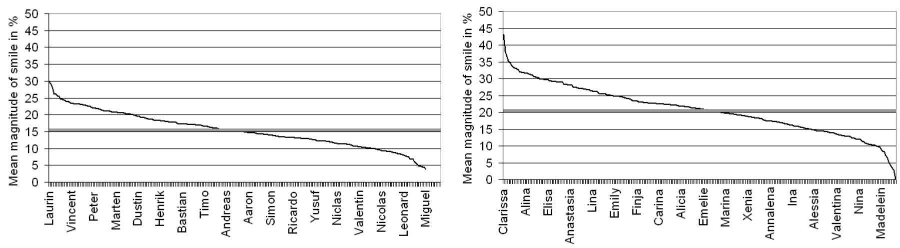


Fig. 3. This figure depicts the distribution of the mean magnitude of smile of the 250 male and 250 female first names. The x-axes just indicate some of the names. The gray bars show the mean values of both genders.

magnitudes of smile for both genders yields, that women are more prone to smile than their male counterparts.

5 Discussion and Conclusion

We demonstrate an approach to derive facial expressions from images with human faces. It integrates various state-of-the-art algorithms for robust model fitting and facial expression estimation. The system is able to automatically conduct a large-scale survey to estimate the influence of persons' first names on their presentation in the Internet.

Currently, this experiment is neither considering the age nor the cultural background of the persons depicted. Cultural background and sense of politeness heavily influence the way people present themselves in the Internet. Furthermore, the same person may be depicted in several images. Especially prominent persons with rare names are depicted more than once and induce unrepresentative information. Furthermore, persons with surnames that also double as popular first names may be depicted several times.

However, the intention of this experiment is merely to demonstrate the opportunities that current Computer Vision technology provides to conduct automated large-scale analyses in the domain of psychological and sociological surveys.

References

1. M. S. Bartlett, J. C. Hager, Paul Ekman, and T. J. Sejnowski. Measuring facial expressions by computer image analysis. *Psychophysiology*, 36:253–263, 1999.
2. Knud Bielefeld. beliebte-vornamen.de, 2008.
3. Tim F. Cootes, G. J. Edwards, and Chris J. Taylor. Active appearance models. In *5th European Conference on Computer Vision*, pages 484–498, Germany, 1998.
4. Paul Ekman. Universals and cultural differences in facial expressions of emotion. In *Nebraska Symposium on Motivation*, pages 207–283, 1972. U of Nebraska Press.
5. Philip G. Erwin. First names and perceptions of physical attractiveness. *The Journal of Psychology*, 127(6):625–631, November 1993.

6. Roland G. Fryer and Steven D. Levitt. The causes and consequences of distinctively black names. *The Quarterly Journal of Economics*, 119(3):767–805, August 2004.
7. Google. Image search, 2008.
8. Robert Hanek. *Fitting Parametric Curve Models to Images Using Local Self-adapting Seperation Criteria*. PhD thesis, Technische Universität München, 2004.
9. Wayne E. Hensley and Barbara A. Spencer. The effect of first names on perceptions of female attractiveness. *Sex Roles*, 12(7-8):723–729, April 1985.
10. Kenneth M. Steele I and Laura E. Smithwick. First names and first impressions: A fragile relationship. *Sex Roles*, 21(7-8):517–523, October 1989.
11. Takeo Kanade, John F. Cohn, and Yingli Tian. Comprehensive database for facial expression analysis. In *Automatic Face and Gesture Recognition*, pp 46–53, 2000.
12. E. D. Lawson. Psychological dimensions of men's names: A semantic differential analysis. *Der Eigenname in Sprache und Gesellschaft*, 3, 1985.
13. P. Michel and R. El Kaliouby. Real time facial expression recognition in video using support vector machines. In *Conf on Multimod Interfaces*, pp 258–264, 2003.
14. R. Quinlan. Learning with continuous classes. In *Proc of the* 5^{th} *Australian Joint Conference on Artificial Intelligence*, pages 343–348, 1992.
15. R. Schweiger, P. Bayerl, and Heiko Neumann. Neural architecture for temporal emotion classification. In *ADS 2004, LNAI 3068*, pages 49–52, 2004.
16. Ying-Li Tian, Takeo Kanade, and Jeffrey F. Cohn. Recognizing action units for facial expression analysis. *PAMI*, 23(2):97–115, February 2001.
17. Paul Viola and Michael J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.
18. Matthias Wimmer. *Model-based Image Interpretation with Application to Facial Expression Recognition*. PhD thesis, Technische Universitat München, 2007.
19. Matthias Wimmer, Freek Stulp, Sylvia Pietzsch, and Bernd Radig. Learning local objective functions for robust face model fitting. *PAMI*, 30(8), 2008.
20. Ian H. Witten and Eibe Frank. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, San Francisco, 2^{nd} edition, 2005.
21. R.L. Zweigenhaft, K.N. Hayes, and C.H. Haagen. The psychological impact of names. *Journal of Social Psychology*, 110:203–210, 1980.

Classification of facial Expressions based on Transient Features

Khadoudja Ghanem[1], Alice caplier[2]
[1]*Laboratory Lire, Mentouri University , Rue Ain El Bey, Constantine, Algeria*
[2]Gipsa Laboratory, 46 avenue Félix Viallet, F-38031 Grenoble, France

gkhadoudja@yahoo.fr *alice.caplier@gipsa-lab.inpg.fr*

Abstract: Over the last decades, the most automatic facial expression recognition methods and systems presented in the literature are based on motions and deformations of permanent facial features and as a post processing they use some transient features to improve some results. In this paper we present a new method which represents the opposed method, it uses firstly transient facial features due to expressions for the expression recognition and then the permanent features for the phase of post processing. Obtained results are compared to the first ones, they reveal that they are almost the same.

Key Words: Facial expression, Classification, Transient features, belief theory.

1. Introduction

Facial expression recognition deals with the classification of an expression into abstract classes postulated by Friesen and Ekman: Joy, Anger, Disgust, Surprise, Sadness and Fear. Facial expressions are generated by contractions of facial muscles which produce deformations on facial features like eyes, eye brows, mouth and skin texture. The widely used approach for expression recognition is that which uses permanent facial features. Indeed, among these studies, some ones were interested by defining witch part of the face is the most pertinent to recognize expressions [1], [2], [3]: half part, bottom part or the whole face. A new idea is proposed in this work which is concerning parts were transient features (furrows and wrinkles) can appear. Several studies in the literature focused on transient features but not to recognize expressions. The presence or absence of furrows in a face can be determined by edge features analysis [4], [5] or by eigen-image analysis [6], [7]. Terzopoulos and Waters [8] detect the nasolabial furrows for driving a face animator, but with artificial markers. Kwon and Lobo [4] detect furrows using snakes to classify pictures of people into different age groups. Ying-li Tian and Takeo Kanade [5] detect horizontal, vertical and diagonal edges using a complex face template, and then they use the Canny edge detector to quantify the amount and orientation of the furrows [9] .

Our point of view is that it is easier and faster to detect if there is or there is no wrinkles on the face than to calculate distances from contours, compare them with those of neutral state and then deduce if they decrease or increase, moreover, it is delicate to detect characteristic points corresponding to corners of eyes, eye brows and mouth to extract distances, however, wrinkles can be easily detected and decide if they are absent or present. Our challenge is to validate this idea and to prove that these features contain enough information to recognize the considered expression.

2. Related work

The proposed approach consists mainly of five steps: segmentation, data extraction, data analysis, classification and post processing. In the segmentation step, two facial images with and without expression are presented to the system and contours of all transient features on different facial regions besides eyes, eyebrows and mouth are located (See Section.2.2).

In the data extraction step, for the classification step, and for each wrinkle region a ratio which correspond to the number of edge pixels of the expressive face with the number of edge pixels of the neutral face, is calculated, and for the post processing step, the angle of the nasolabial furrows if exists

besides, some distances which correspond to the degree of opening or closing eyes, and the raising or lowering of the eye brow are calculated to know if they increase or decrease.
In the data analysis step, we associate each facial wrinkle region with a state “present” or “absent”, then we characterize each wrinkle region by a combination of the corresponding expressions. This process is done after a training step(See Section. 2.3), were we determine which wrinkle corresponds to which expression.
In the classification step, the Transferable Belief Model (TBM) is applied to recognize the facial expression.
Finally, the post processing step (See Section. 2.5), refine the obtained result from the classification step by reducing the doubt between possible expressions. In this step, some distances are mapped to states that encode how much a given distance differs from its corresponding value in the Neutral state (Decrease or Increase), and if exists, the angle of the nasolabial furrow is compared to a threshold obtained from another study on the nasolabial furrows which determine thresholds differing between Disgust and Joy.
2.1. Study of presence of transient facial features: Different kinds of transient features occur with facial expressions. Sometimes on one region and in other times on several regions, this is due to expression intensity. Here nine regions are studied which are: The forehead, the regions surrounding eyes and mouth, nasal root, on the nasolabial regions and finally on the chin. Wrinkles on and under lids can occur too but they are not considered in this work because for the first ones, they are generally confused by the lid or the eye brow, and for the second ones their appearance is always associated with nasolabial furrows.

2.2 Detection of facial transient features:
In this paper we use the Canny edge detector and the threshold of the interest region method to detect the presence or absence of transient features in the nine regions on a face compared to the neutral face. The location of the regions of interest are determined by the location of characteristic points of permanent features.
We compare the number of edge pixels in the wrinkle areas of the expressive face with the number of edge pixels of the neutral face. If the ratio is larger than a threshold, the furrows are determined to be present. Otherwise they are absent.

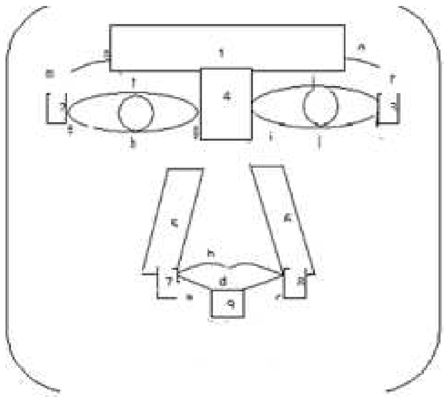
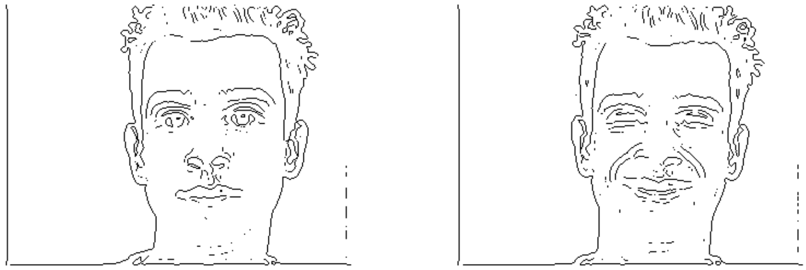


Fig. 1 : The nine regions corresponding to transient facial features



-a- -b-

Fig.2 : (a):Face without transient features ;(b): Face with transient features on the nasolabial regions

2.3 Results of presence or absence of transient features with different facial expression databases:

To generalize the presence or the absence of all facial transient features with the six universal expressions, we trained our program on 90 images coming from Hammal_Caplier database [10], on 462 images coming from EEbase database [11] and on 144 images coming from DAfex database [12]. Using the Canny edge detector, we detect all transient facial features appearing with each facial expression. We calculate the rates of presence of wrinkles on each region of the nine regions corresponding to transient regions, and at the end we transform these results into a logical table (see Table 1). This table gives the logical rules for each interest region:”1” if for the considered facial expression, wrinkles appear in the corresponding interest region and ”0” otherwise. This procedure ensures a validation of the presence or absence of wrinkles with each class of facial expressions across different databases.

Transient features	JOY	SURPRISE	DISGUST	ANGER	SADNESS	FEAR
Chin(1 region)	(0)	(0)	(0)	(1U0)	(1U0)	(1U0)
Mouthcorners (2 Left & right regions)	(0)	(0)	(0)	(0)	(1U0)	(1U0)
Nasolabial (2 Left & right)	(1U0)	(0)	(1U0)	(1U0)	(1U0)	(1U0)
Eyes corners (2 left & right eyes)	(1U0)	(0)	(1U0)	(1U0)	(0)	(0)
Nasal root (1 region)	(0)	(0)	(1U0)	(1U0)	(1U0)	(1U0)
Forehead (1 region)	(0)	(1U0)	(0)	(1U0)	(1U0)	(1U0)

Table 1 : Presence or absence of transient features on each region for each facial expression.

On table 1, columns represent the six universal expressions and the lines represent the nine regions where wrinkles can appear. For example with surprise we can get wrinkles only on the front head and with joy we can get wrinkles on nasolabial regions and on eye corners regions.

2.4 Data Analysis and classification using the transferable belief model:

After the extraction data step, a conversion of numerical data (Calculated ratios) to symbolic data is done, if the ratio is larger than a threshold, we associate the state “1 or present” to the concerned wrinkle region, otherwise the state “0 or absent” is associated to this region. After that we characterize each wrinkle region by a combination of the corresponding expressions using table 1(See example in section 2.4.1).

As we have several extracted data, all these data will be fused. The Transferable Belief Model is a powerful fusion process[13], [14], [15], [16]. Its salient character is the powerful combination operator that allows the integration of information from different sensors. In order to make the decision about the facial expression, these sensors are combined to take into account all the available information.

This fusion method is well suited for the problem of facial expressions classification: this model facilitates the integration of a priori knowledge and it can deal with uncertain and imprecise data which could be the case with data obtained from automatic segmentation algorithms. It is well adapted too when the learning set is incomplete, due to a lack of learning data on one or more regions(absence of wrinkles). In addition it is able to model intrinsic doubt which can occur between facial expressions in the recognition process (see Figure 3). The classification of any facial expression into a single emotion category is not realistic because human expressions are variable according to the individual. Moreover, sometimes the emotion is not clearly expressed and then cannot be directly recognized. For all these reasons , we have choose this method which is well adapted for such conditions.

Fig 3 : Example of doubt between Surprise and Fear

2.4.1 Belief Theory principle [17], [18]:

The Belief Theory is a generalization of the probability theory [19]. It has been introduced by Dempster & Shafer and then by Smets. It requires the definition of a set Ω = {E1,E2,…,EN} of N exclusive and exhaustive assumptions. We also consider the power set 2^{Ω} that denotes the set of all subsets of Ω. To each element A of 2^{Ω} is associated an elementary piece of evidence *m(A)* which indicates all confidence that one can have in this proposal. The function *m* is defined as:

$$m: 2^{\Omega} \rightarrow [0,1] \quad A \rightarrow m(A) \quad \text{where : } \Sigma m(A)=1 \tag{1}$$

As we have several sources of information, we have to take into account all the available information. The global evidence is obtained using the rule of conjunctive combination or orthogonal sum. In the case of two assumptions, the orthogonal sum is defined in the following way:

$$m=m1+m2 \quad m(A)=\sum m1(B)+m2(C) \tag{2}$$

This could give propositions whose number of elements is lower than the initial ones and to associate them a piece of evidence. The final evidence is thus more accurate.

In our application, the first sensor Ω corresponds to the six facial expressions: *joy*, *surprise*, *disgust, Anger, sadness and fear* ; 2^{Ω} corresponds to single expressions or combinations of expressions and A is one of its elements. The second sensor Ω' corresponds to the two states taken by each wrinkle region : *present* and *absent*.

In the other hand, and for each region where wrinkles can appear, we associate the piece of evidence "1" to the region if wrinkles are present in this region and associate the piece of evidence "0" to this region if the wrinkles are absent. This piece of evidence corresponds to the evidence of all considered expressions where wrinkles appear in these regions.

To be more explicit we propose this example: we suppose that we have two facial regions were wrinkles appear, nasal_root and the nasolabial regions. The associated states and mass of evidence are:

- m(nasal-root)(present)=1
- m(naso)(present)=1

The characterization of each wrinkle region by a combination of the corresponding expressions using table 1 is:

- m(nasal-root)=m(Disgust OR Anger OR Sadness OR Fear)=1
- m(naso)=m(Joy OR Disgust OR Anger OR Sadness OR Fear)=1

The fusion process is then :

⇨ m(nasal-root_naso)(Disgust OR Anger OR Sadness OR Fear) = m(Disgust OR Anger OR Sadness OR Fear). m(Joy OR disgust OR Anger OR Sadness OR Fear) = 1

It means that if we have two regions (nasal-root and nasolobial) were wrinkles appear, the corresponding expression can be one of the four expressions (Anger, Disgust, Sadness or Fear). Generally, final result presents doubt between several expressions, so, we proceed to a post processing step to reduce this doubt. The post processing corresponds to a study of some permanent features.

2.5 Post processing: Study of permanent features

Different facial expression descriptions were proposed in the MPEG-4 literature [20] [21], [22] , [23], [24], [25], for example, in one description, anger is defined by tightening lips, and in another one the mouth is stretched and opened(fig.4).

-1- -2- -3- -4-

Fig.4 : Different descriptions of Anger from left to Wright :
(1)Eyes are slightly opened and mouth is opened;
(2,3)eyes are normally opened and lips tightened ;
(4)Eyes are strongly opened and mouth too.

In order to release the potential differences between the six universal expressions, we propose a new description which takes into consideration all precedent descriptions. As instance, and for the same expression(anger), the new description define the mentioned expression by tightening the lips OR opening the mouth.

	Distance between lids	Distance between eye and eye brow	Distance between the mouth corners	Distance between the top and bottom of the higher and lower lips	Distance between eye corner and mouth corner
Joy	Increase or decrease	Increase, relaxe or decrease	increase	Relaxe or increase	decrease
Surprise	increase	increase	Relaxe or decrease	increase	Relaxe or increase
Disgust	Decrease	Decrease	Increase, relaxe or decrease	increase	Increase, relaxe or decrease
Anger	Decrease or increase	Decrease	Relaxe or decrease	Increase, relaxe or decrease	Relaxe or increase

Sadness	Decrease	increase	Relaxe or increase	Relaxe or increase	Relaxe or decrease
Fear	Relaxe or increase	Relaxe or increase	Increase, relaxe or decrease	Relaxe or increase	Increase, relaxe or decrease

Table 2: New proposition of descriptions of the universal expressions.

After that we compare every two expressions to deduce differences in the orientation of the deformation of each permanent feature between both. In the case of joy, the difference with surprise and anger is: with these two expressions, the mouth is opened vertically but with joy it is opened horizontally. For the difference of joy with the other expressions, deformations of permanent features can evolve in the same direction, for example, in the case of Joy and Disgust, the mouth can be opened horizontally (like with joy), or can be opened vertically.

In the same way we compare each expression with all the other ones and deduce the differences (see table 3). For the case where there is no difference between two expressions like between surprise and fear, the doubt cannot be removed. Table 3 is deduced from table 2.

Expressions	Differences
Joy versus surprise	For the first expression, distance between mouth corners increases; for the second expression, distance decreases or relaxes
Joy versus anger	For the first expression, distance between mouth corners increases; for the second expression, distance decreases or relaxes
Surprise Versus Disgust	For the first expression, distance between lids increases; for the second expression, distance decreases The same thing with Distance between eye and eye brow
Surprise versus Anger	For the first expression, distance between eye and eye brow increases; for the second expression, distance decreases
Surprise versus Sadness	For the first expression, distance between lids increases; for the second expression, distance decreases
Disgust versus Sadness	For the first expression, distance between eye and eye brow decreases; for the second expression, distance increases
Disgust versus Fear	For the first expression, distance between lids decreases; for the second expression, distance increases or relaxes The same thing with distance between eye and eye brow
Anger versus Sadness	For the first expression, distance between eye and eye brow decreases; for the second expression, the distance increases
Anger versus Fear	For the first expression, distance between eye and eye brow decreases; for the second expression, distance increases or relaxes
Sadness versus Fear	For the first expression, distance between lids decreases; for the second expression, distance increases or relaxes

Table 3: The potential Differences between universal expressions.

To be more explicit, we consider the result from the study of presence of transient features for an actor from the Dafex database (the actor presents Anger where the distance between the interior eye corner and inner brow decreases), the actor presents wrinkles on the nasolabial region and on the nasal root. The analysis result is a doubt between disgust, anger, sadness and fear. The post processing stage, allows us to eliminate sadness because the difference between sadness and anger(Table 3), is in the distance between the eye and the eye brow which decreases with anger and increases with sadness. In another hand we can eliminate fear because the difference between fear and disgust is in the distance

between the eye and the eye brow which decreases with disgust and increases with fear. As there is no difference (in facial permanent deformations) between disgust and anger the doubt between these two expressions cannot be removed and the last result is Disgust OR Anger instead of :
Disgust OR Anger OR Fear OR Sadness.

We proceed in the same way with all actors of the Dafex database, so, we detect all transient features on all images of the database(78 subjects have transient features), for each subject, we proceed to a data fusion using the belief theory, data correspond to the presence of transient features in different regions of the face, then we apply a post processing to obtained results in order to reduce doubt between expressions. At the end, we sum all similar results and compute rates according to the total number of subjects for each facial expression. Final results are presented in table 4 (moy=Expression Intensity is medium; max=Expression intensity is high).

EXPERT / SYSTEM	Joy_ moy	Joy_ max	Dis moy	Dis max	Ang moy	Ang_ max	Sad moy	Sad max	Fea_ moy	Fea_ max	Sur moy	Sur max
Joy												
Disgust												
Anger												
Sadness							100	100				
Fear									66,67 %	42,86 %		
Surprise												
JoyORDis	87,5 %	87,5 %										
AngORDis			100	100	100	85,7 %						
FearORSurp									33,33 %	57,14 %	100	100
Error	12,5 %	12,5 %				14,3 %						
Total	100	100	100	100	100	100	100	100	100	1OO	100	100

Table 4: Primary expression classification based on presence of transient features with post processing on permanent features.

From this table we can see that joy is generally confused with disgust, disgust is generally confused with anger and surprise is generally confused with fear. Sadness is the only expression which is not confused with any other expression.

An information conveyed by the nasolabial furrows which is the angle formed between the adjusted line representing the nasolabial furrow and the horizontal one, connecting the two corners mouth, allows to improve the obtained rates in the case of the doubt between joy and disgust, especially when the intensity expression is high, because the mentioned angle in case of joy is lower than the angle formed in case of disgust.

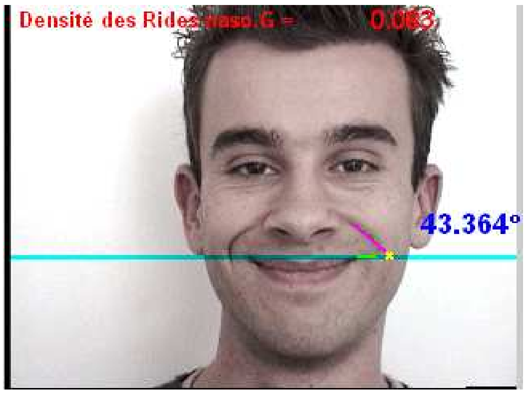


Fig.5 Detection of nasolabial furrows and the computation of the nasolabial angle

So, the new rates have changed for joy. When the expression intensity is medium, the rate representing doubt between Joy and Disgust decreases from 87,5% to 37,5%, in favor of the recognition of joy expression (50%). When the intensity expression is high the whole rate becomes the rate of recognition of joy expression.

2.6 Comparison with other classification systems:

To evaluate the performances of our approach, we compare the current system (based on transient features) with that of Hammal et al. [14] (based on permanent features).

	Hammal et al research/ CKE Database	Our research/ Dafex Database
Joy	64,51%	50% if intensity = Medium 87,5% if intensity = high
JoyORDisgust	32.27%	37.5% if intensity =medium
SurpriseORFear	84%	100%

Table 5: Comparison of recognition rates between our system and another system [14]

Comparisons are summarized in table 5. Hammal et al system gives rates of recognition of three expressions :Joy, Disgust and Surprise. We have compared results of these three expressions of this system with results of the same expressions obtained from our system. For joy expression, the classification rates are very comparable, when intensity expression is high, best performance are given by our classification system. In another hand, the both systems introduce two sources of confusion the first one is between Joy and Disgust, and the second one between Surprise and Fear. Classifications rates of JoyORDisgust and SurpriseORFear are almost the same.

These results provided converging evidence for the similarity in using either transient facial features or permanent facial features, the obtained classification rates are almost the same. Another source of confusion is introduced by our system classification : doubt between Disgust and Anger, this source does not appear with the other system, because anger is not evaluated in the referenced system.

3. Conclusion

Automatically recognizing expressions based on the presence of facial transient features is a new research orientation. Obtained results have proved that this approach can be an accepted foundation for recognizing facial expressions. Even if in the most cases there is a doubt between two expressions at most, it is preferable to keep the doubt between these two expressions instead of taking the risk of choosing the wrong one. This is why we use the transferable belief theory to model the doubt and for fusion of data.

Our comparison indicates that a permanent feature based method performs just as well as the transient feature based method. We feel that we have to combine both approaches to achieve optimal performances.

References:

[1] Boucher J.D. and Ekman P. Facial areas and emotional information. Journal of Communication, 25 no. 2:21–29, 1975.

[2] Bassili J. N. Facial motion in the perception of faces and of emotional expression. Experimental Psychology - Human Perception and Performance, 4 no.3:373–379, 1978.

[3] Gouta K. and Miyamoto M. Facial areas and emotional information. Japanese journal of psychology, 71 no. 3:211–218, 2000.

[4] Y. Kwon and N.Lobo, « Age classification from facial images » proc. IEEE Conf. Computer vision and Pattern Recognition, p. 762-767 , 1994

[5] J.JLien, T.Kanade,J.F Cohn , « Detection tracking, and classification of Action Units in Facial expression », J.Robotics and autonomous system, vol.31, p. 131-146, 2000.

[6] M.Kirby et L.Sirovich, « Application of the K-L Procedure for the characterization of Human Faces », IEEE Trans. Pattern analysis and machine intelligence, vol 12, N°1, p.103-108, jan. 1990.

[7] M.Turk et A.Pentland, « Face recognition using eigenfaces », Proc. IEEE Conf. Computer vision, p. 727-732 , 1990.

[8] D.Terzopoulos et K.Waters, « Analysis of facial images using physical and anatomical models », proc. IEEE Intl'Conf. Computer vision and Pattern Recognition, p. 762-767 , 1994

[9] Y.Li Tian, T.Kanade et J.F.Cohn « Recognizing action units for facial expression analysis",IEEE Trans. Pattern analysis and machine intelligence, vol 23, N°2, p.97-114, feb.2001.

[10] Hammal_Caplier data base :
http://www.lis.inpg.fr/pages_perso/caplier/english/emotionnelle.html.en/emotionnelle_2.html.en.html

[11] EEBase Database: http://www.cs.cmu.edu/afs/cs/project/cil/ftp/html/v-images.html

[12] Dafex Database: http://tcc.itc.it/research.i3p/dafex/index.html

[13] V. Girondel, A. Caplier, L. Bonnaud, and M. Rombaut. "Belief theory-based classifiers comparison for static human body postures recognition in video". Int. Jour. of Signal Processing, 2(1) :29–33, 2005.

[14] T. Denoeux and Ph. Smets. "Classification using belief functions : The relationship between the case-based and model-based approaches". IEEE Trans. on Systems, Man and Cybernetics, 2006.

[15] D. Mercier. "information Fusion for automatic recognition of postal addresses with belief functions theory" PhD thesis, University of Technologie of Compiegne, December 2006.

[16] E. Ramasso, C. Panagiotakis, M. Rombaut, and D. Pellerin. "Human action recognition in videos based on the transferable belief model - application to athletics jumps". Pattern Analysis and Applications Journal, 2007.

[17] A. Dempster. "A generalization of Bayesian inference". Journal of the royal statistical society, Vol30, ages 205-245, 1968.

[18] G. Shafer. "A Mathematical Theory of Evidence". Vo.l2702. Princeton,1976.

[19] PH. Smets "Data Fusion in the Transferable Belief Model." Proc. 3rd Intern. Conf. Information Fusion, Paris, France Pages 21-33. 2000.

[20] M. Tekalp M. "Face and 2d mesh animation in mpeg-4". Tutorial Issue on the MPEG-4 Standard, Image Communication Journal, Elsevier, 1999.

[21] Hammal Z., Caplier A. and Rombaut M. "A fusion process based on belief theory classification of facial basic emotions". Proc. the 8th International Conference on Information fusion, (ISIF), Philadelphia, PA, USA, 2005.

[22] J.M Carroll, J.Russell," Facial Expression in hollywood's Portrayal of Emotion",J.Personality and social psychology, vol.72, p.164-176, 1997.

[23] I.Eibel-Eihesfeldt, Human Ethology. New work: Aldine de Gruvter,1989.

[24] M. Pardàs. Extraction and Tracking of the Eyelids, Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 4, pp. 2357-2360, Istambul, Turkey, June 2000.

[25] N. Tsapatsoulis, K. Karpouzis, G. Stamou, F. Piat, S.A Kollias, "A fuzzy system for emotion classification based on the MPEG-4 facial definition parameter set", Proc. 10th European Signal Processing Conference, Tampere, Finlande, September 2000.

About Emergence of Emotional Behavior - An Experiment using the Public Goods Game Scenario

Dirk M. Reichardt

BA Stuttgart - University of Cooperative Education
D-70180 Stuttgart, Germany
reichardt@ba-stuttgart.de

Abstract: Our experiment tries to mimic the development of emotions and their influence on decision making. The public goods game scenario was chosen for simplification of the decision situation. The approach combines a model of emotion and evolutionary computing techniques to see how a population of virtual agent generations evolves. Our virtual emotional agents integrate a model of emotion based upon the cognitive appraisal theory by Ortony, Clore and Collins combined with the NEO-FFI ("Big 5") model of personality and a mood component. First results show the effects of introducing an emotional agent as discrete strategy and the evolution of agent configurations.

Introduction

Emotion is considered a key to human social behavior and explains social phenomena like the emergence of costly punishment. In previous work we developed an agent which shows emotional behavior. Would virtual emotional agents show aspects of the behavior of human societies? Instead of trying to implement the required behavior which comes closest to the observed human social behavior, the idea of the experiment described here is to let the agent society evolve. The personality configurations and the emotional behavior should emerge from interactions in the experiment. The motivation for the experiment is twofold. One idea is to test the emotional model for operational capability and also plausibility considering a society of humans and of artificial emotional agents as basis for comparison (see also [3]). Is it possible to determine which group consists of humans and which group does not? The second and not less important motivation is the curiosity to see how emotional agent societies evolve.

The public goods game with punishment option is a well suited small scenario in which affect-based decisions are shown in human players [1]. Based upon previous work, we adopted the basic model of Ortony, Clore and Collins [2] enhanced by a concept of mood and by a personality model.

In this paper we present the architecture of our emotional agent and first experimental results in which virtual agent societies are simulated.

The Scenario

Punishment is used to enforce cooperation. Interesting effects on human behavior can be observed in the public goods game scenario. A short scenario description of the public goods game: n players form a group, each player has the choice to invest the amount c Euro in a public project which is bearing r % interest. The sum of investments of all n players is distributed to the players in equal shares plus interest. After the publication of the investments each player can decide whether to punish another player or not. The punishment cost for the punisher is p and punishment results in $f \cdot p$ deduction from the account of the punished player.

The discrete version of the game – as used in studies of emergence of behavioral strategies (see [4]) - distinguishes between the roles of the punisher, the cooperator and the defector. The punisher and the cooperator invest whereas the defector does not contribute to the public project. If n_p is the number of punishers in the population, n_c the number of cooperators and n_d the number of defectors, the payoffs can be computed as follows. The public good value is $(r+1)(n_p+n_c)c/(n_p+n_c+n_d)$. The payoff for the cooperators is determined by subtracting the investment c. For the defectors we need to subtract their punishment which is fpn_p. The punisher's payoff is reduced by c and by the punishment cost which is pn_d. In the iterated discrete game, the players are allowed to change the strategy from time to time. This is simulated by a strategy change which contains an element of chance and prefers strategies which show good performance. Moreover, a 'mutation' element is introduced allowing (very rarely) the re-introduction of a strategy that probably already disappeared. As shown in [4] the iterated public goods game leads to a predominance of the defectors. This is changed by giving the participants the freedom to choose whether to participate in the game as such or not. The nonparticipant gets the amount of c. In case of a dominance of defectors, it is likely that c is larger than the payoff the player would gain when participating. As an effect, not participating is better in this situation and the number of defectors in the game decreases. Instead, the punishers get a chance to get back into the game and it is shown that they become dominant (see [4]). In contrast, a system without punishers is unstable and the dominance is transient and switches from defectors to nonparticipants to cooperators back to the defectors.

In our experiments we simulated two different settings of the scenario described above. First tests are done in a discrete version using fixed strategies as alternative agent configurations together with an emotional agent configuration. In a second experiment we simulated emotional agents only and observed the evolution of personality configurations and emotional behavior.

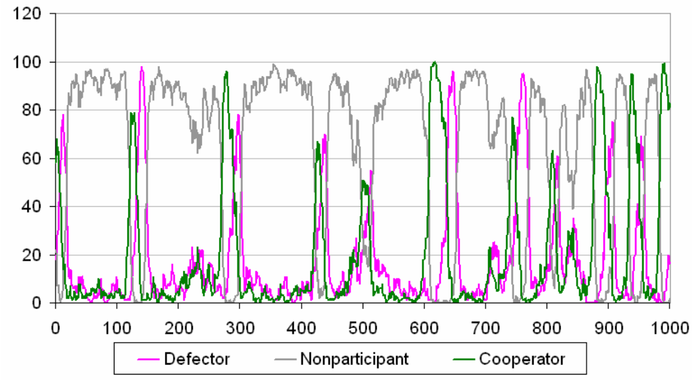


Fig. 1. Cyclic behavior of the population without punishers as described in [4]. The behavior is produced by our simulation tool which allows the integration of a virtual emotional agent.

The Model of Emotion

The emotion model by Ortony, Clore and Collins (OCC) [2] is used in our agent as basic framework. It distinguishes 22 emotion categories and defines an interface to the application by intensity variables. The main task in order to use the model in an application is to adapt the intensity values to the application scenario - in our case to the public goods game with punishment option. The mapping of intensity variables to our scenario is described in greater detail in [5][6]. As an example, the intensity variables likelihood and desirability define the main influence on the emotions joy and distress. The expected payments and punishments are computed based upon a locally stored action history for every other agent. This forms the rational basis of the agent.

In our approach we introduced a concept of mood. The idea is to have a mid-term multidimensional accumulator which serves as an affect filter. The computed emotion intensity is biased by the mood such that the same stimuli cause different emotional reactions depending on the mood a person is in. Each emotional experience modifies the mood which shows an inert behaviour as compared to emotions.

To make the two concepts of emotion and mood compatible we use the pleasure-arousal-dominance (PAD) space (see [7][8]). First, the mood is shifted in the direction of the current emotion. This can be described by the metaphor of the attraction of two masses with mood being the much bigger one. In the other direction, mood does not change the *quality* but in the *elicitation* (strength) of the emotion. The elicitation value is reduced depending on the distance between emotion and mood in PAD-space.

As additional concept, the personality of an agent can be seen as a bias for the emotion generation. The term personality describes the factors that determine the character of a human individual. The "big 5" model of personality is the most popular model nowadays [9]. It consists of the traits neuroticism, extraversion, openness, agreeableness and conscientiousness. We integrated this model in our virtual emotional agent such that personality parameters influence the emotion elicitation.

The elicitor functions contain the intensity variables as parameters. As an example, high values on neuroticism increase the elicitation level of negative emotions. Another influence is set directly to intensity variables. For example, high values on openness increase the sense-of-reality intensity value. Personality also influences the dynamics of mood changes. Additionally we adopted the feature of a long term mood attractor which is defined by the personality [8].

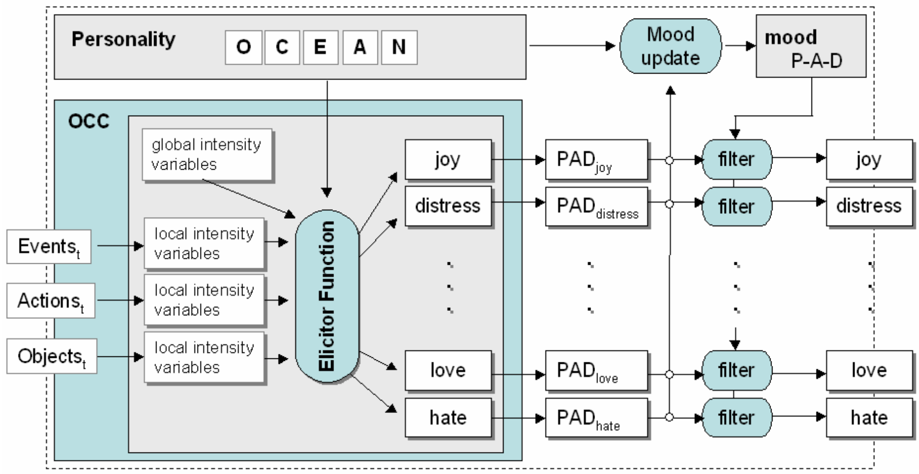


Fig. 2. A system overview of the emotion system of the virtual agent. It shows the influences of the personality model and the mood concept.

The emotional agent in our simulation is configured by setting the personality traits and two further specific parameters called *goalpreference* and *moraldisposition*. The moral disposition determines whether and to what extent the agent shows shame or pride. The goal preference parameter is a simplified approach to the rational vs. emotional decision making. There are only two decisions in the scenario: contribution and punishment. In the simulation these are binary decisions (in contrast to earlier publications which use different settings). Decisions are always prepared by rational reasoning. The only reason to punish is given if a threshold of the elicitation of the emotion *anger* is surpassed. In case of the contribution decision, the system is designed differently. We pre-calculate the potential action and reactions of the others using a log history. For each potential future state, the system computes the emotional state this would lead to. With this evaluation we get an emotional decision. If emotional and rational decision differ, the *goal preference* setting determines which one to take.

First Simulation Results

In the first simulation scenario we introduced the virtual emotional agent as an agent strategy in a discrete game (see [4]). The virtual emotional agents are initialized by personality values as well as by the two parameters goal preference and moral disposition. In order to get plausible values for the parameters, we performed a personality test with human test persons (anonymously). The figure below shows a test run in which we fixed the number of emotional agents. These agents replaced cooperators *and* punishers from the simulation scenario we ran before since emotional agents can show both kinds of behavior. It can be observed that in the presence of emotional agents the number of non-participants and defectors is almost stable. As can be seen in the test run, there are some disturbances - which can be explained by the virtual agent behavior.

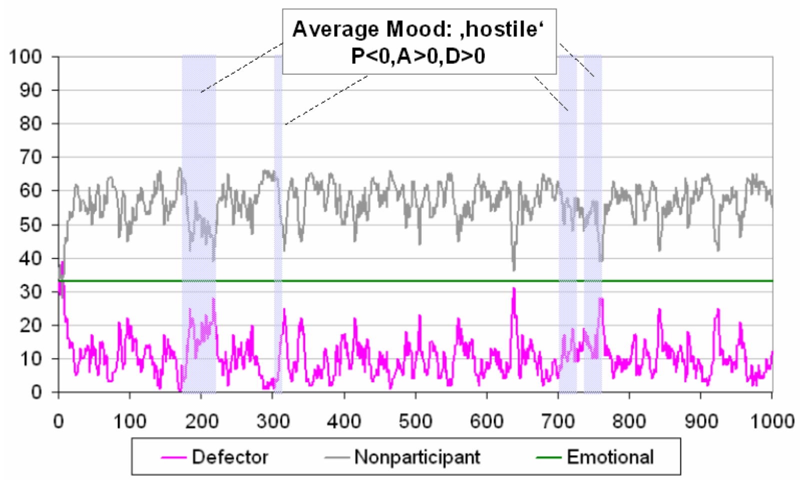


Fig. 3. The figure shows the number of agents (strategies) over played rounds of a test run with a fixed size subgroup of emotional agents. Small effects (low significance) on the population development are visible in the phases in which the average mood turns to 'hostile'.

Due to the high number of defectors the payoff is rather low, therefore the mood of the agents becomes negative. In the marked regions, the average mood is p<0, a>0, d>0 ('hostile'). This fits the parameters of the emotion ‘anger’ very well, therefore the highest elicitation values for anger can be seen in these regions. This leads to punishment decisions which make defecting less attractive and consequently their number decreases.

This first simulation does not include any consequence for the punishing emotional agents. Their number is not reduced due to the loss of resources. This is changed in the settings of our second simulation. Here we use virtual emotional agents only. The agent population (size = 100) is initialized by random choice of personality trait values. After ten consecutive games, the population is updated using a probabilistic tournament selection (size 2). The agents with a better performance have a higher chance to survive. Moreover, mutations are introduced such that new configurations of agents (random choice of traits) are injected in the population (mutation probability 1/100). First simulation runs show a very quick stabilization of the observed behavior: no contribution and no punishment.

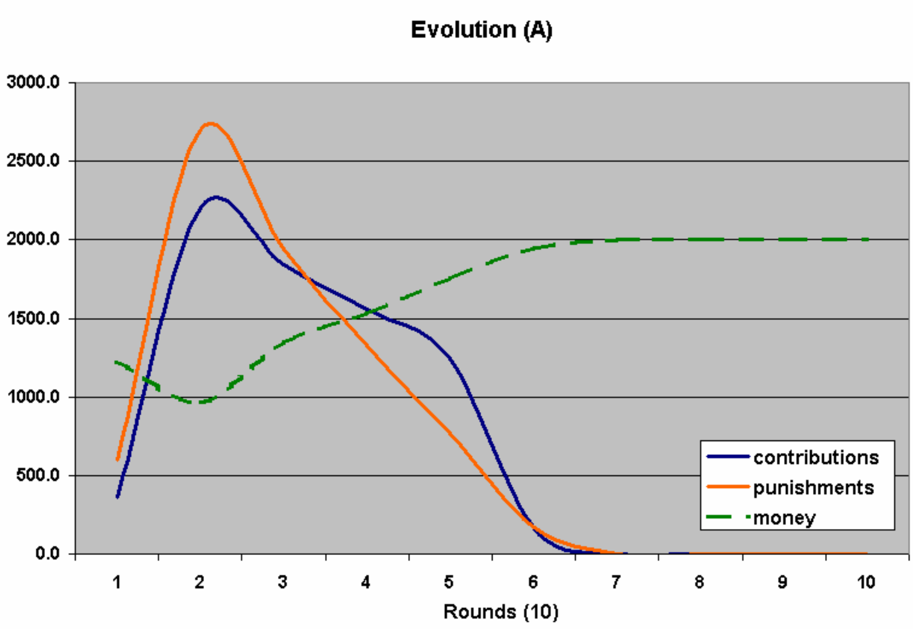


Fig. 4. A test run which shows quick stabilization on the observed behavior *defect and don't punish*. The y-axis shows monetary contributions and outcomes of the 100 participating agents per round.

The mutations lead to short term changes which very quickly lead to the uniform defection behavior again. Another observation is the following: the agent behavior always couples contribution and punishment and the negative effect of punishment on the own resources makes the agents disappear in the evolution process.

Conclusion and Future Work

What can be concluded by now? We can see that emotional agents disappear in the evolutionary setting – but what is the reason? We will have to put some work into other potential configurations of emotional decision making of which we just tested one approach. Besides an emotional model which might just not fit well enough we may consider the goal itself and with that the evaluation basis for the evolutionary algorithm: we evaluated the monetary outcome only! What if we introduce a concept like the compliance of the predominant mood with the personality inherent mood goal state? In other words: it is not always the money which motivates humans, sometimes other aspects count! Future work will try to address these aspects.

References

[1] Fehr, E., Gächter, S., "Altruistic Punishment in Humans", Nature, Volume 415, pp. 137–140., 2002

[2] Ortony, A., Clore, G., Collins, A., "The Cognitive Structure of Emotions", Cambridge University Press, 1988

[3] Reichardt, D., "A definition approach for an emotional Turing Test", in Proceedings of the Second International Conference on Affective Computing and Intelligent Interaction (ACII), Lisbon, Portugal, LNCS 4738. Springer Verlag, 2007

[4] Hauert, C., Traulsen,A., Brandt,H. , Nowak, M. , Sigmund,K., "Via Freedom to Coercion: The Emergence of Costly Punishment", in: Science 316, pp.1905-1907,2007; ISSN 0036-8075

[5] Reichardt, D.," Will Artificial Emotional Agents Show Altruistic Punishment In The Public Goods Game?", in Proceedings of the 1st Workshop Emotion and Computing – Current Research and Future Impact", Reichardt, D., Levi, P., Meyer, J.-J. Chr. (Editors), 29th Annual German Conference on Artificial Intelligence, Bremen, 2006, ISBN 3-88722-664-X

[6] Reichardt, D., "Interpretation of Intensity Variables for an Emotional Agent in the Public Goods Game", in: Proceedings of the 2nd Workshop Emotion and Computing – Current Research an Future Impact, Dirk Reichardt, Paul Levi (Editors), Osnabrück, Germany, September 10th, 2007, ISSN 1865-6374

[7] Mehrabian A., "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament", in: Current Psychology, vol. 14, 1996, pp. 261-292.

[8] Mehrabian A., "Analysis of the Big-five Personality Factors in Terms of the PAD Temperament Model", Australian Journal of Psychology, vol. 48, 2, 1996, pp.86-92.

[9] Gebhard P.,Kipp, K.H., "Are Computer-generated Emotions and Moods plausible to Humans?", in: Proceedings of the 6th International Conference on Intelligent Virtual Agents (IVA'06), pp. 343-356, Marina Del Rey, USA, 2006

Towards a platform for the education in emotion modeling based on virtual environments

Michal Bída, Cyril Brom

Charles University in Prague, Faculty of Mathematics and Physics
Dept. of Software and Computer Science Education, Prague, Czech Republic.
michal.bida@seznam.cz, brom@ksvi.mff.cuni.cz

Abstract. The emotion modeling is a wide, multifaceted theme attracting many scientists. Many different projects focus on different aspects of emotions. What has not been studied in a greater depth is the education in the field of emotion modeling itself. We believe that the time has come to establish a platform that would enable learn-by-example education in this field. This position paper presents our aim at creating a software toolkit facilitating the basic education of event appraisal and emotion modeling in virtual environments. This toolkit is presently being built upon the platform Pogamut 2 used for education of undergraduate students in the field of virtual agents control.

Keywords: Emotions, virtual agents, education.

1 Introduction

Emotions play an important role in human behavior [1]. Due to their ambiguous nature, there are number of ways they can be studied. In recent years there has been a growing interest in emotion modeling in virtual environments. Emotions serve a number of purposes here – from increasing believability or performance of virtual agents to the simulation of some real world scenarios for learning purposes [2]. The role of emotions in virtual agents is not clear-cut. Emotions can affect decision making [3], expressive part of agents' behavior [4] or just serve the agent as another source of information when interacting with humans [5]. So far just a little attention has been given to the education in the field of emotion modeling itself. Especially – as far as we know – there is no platform that would facilitate the education in this field.

In our previous work we have created an educational platform Pogamut 2 [6]. This platform is fully developed and is being used as a tool for education of undergraduate computer science students at Charles University in Prague in the field of modeling behavior of virtual agents. Pogamut 2 enables simple way of prototyping virtual agents in the virtual world of 3D action game Unreal Tournament 2004 (UT04) [7]. In UT04 the player can see the 3D environment through the eyes of the avatar he or she is currently controlling. The 3D environment of UT04 can be very realistic (Fig. 1). Although UT04 is an action game, its environment can be easily exploited for "peaceful" scenarios.

The established university course that uses platform Pogamut 2 concerns also problematics of emotion modeling. So far the education in this field was just theoretical because the Pogamut 2 does not provide any support for emotions. Consequently, the students were unable to take a grasp of the problematics. This fact and the absence of a platform facilitating education of emotion modeling in general motivated us to extend Pogamut 2 with emotion aspect.

Our general plan is to provide an interactive tool that will complement theoretical education in the field of virtual agents and emotion modeling. The target audience are students of computer science interested in artificial intelligence, virtual agents and/or artificial emotions. Educational goals are to provide examples in interactive scenarios covering the following topics: roles of emotions for virtual agents (emotions and expression, emotions as information, emotions and control) and OCC like models (general overview of the mechanisms used, event appraisal). OCC [8] is a cognitive theory of emotions proposed by Ortony, Clore and Collins that provides a popular descriptive emotion model.

We plan to evaluate the project in the already established course that is taking place on Charles University in Prague, Faculty of Mathematics and Physics.

In this position paper, we present the emotion model we have started to develop, including example scenarios intended to use for education (Section 2). We also discuss our approach and open issues of the project (Section 3).

Fig. 1. Environment of the game UT04. UT04 features various locations from open locations (forests, hills, grass fields) to closed ones (old factories, castles, space stations with low gravity, etc.). Copyright Epic Games 2008.

2 Emotions in Pogamut 2 (work in progress)

Pogamut 2 is a software platform for an easy prototyping of virtual agents living in a virtual environment. The aim of the platform was especially to provide the Integrated Development Environment (IDE) (Fig. 2). The platform facilitates the development of virtual agents by offering the IDE, which provides tools for controlling the agent, visualization of agents' inner state and debugging.

The IDE capitalizes on NetBeans. As a virtual environment, 3D action game Unreal Tournament 2004 (UT04) is used. GameBots (GB) [9], debugged and extended by additional features, are utilized to connect UT and IDE via network text protocol.

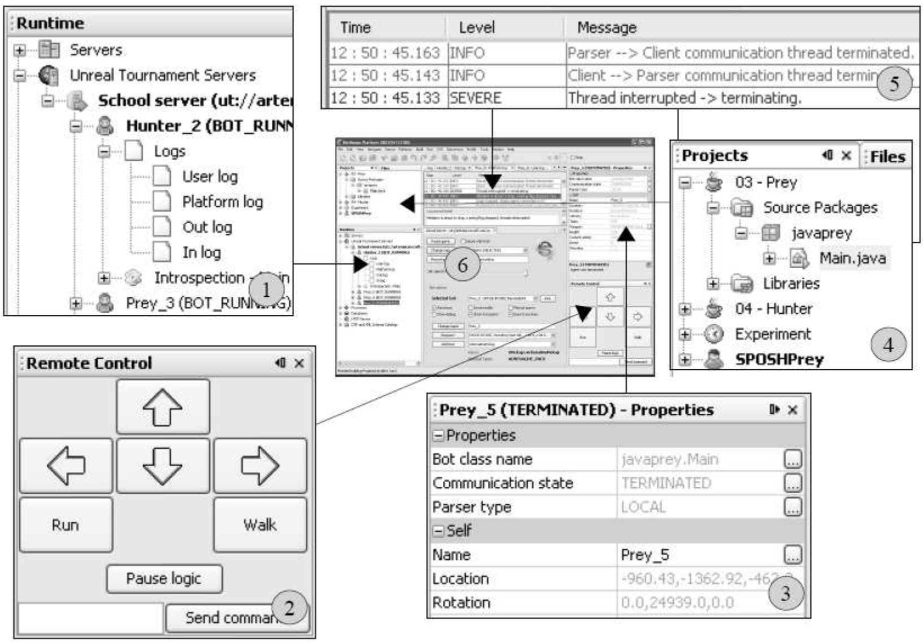


Fig. 2. Pogamut 2 IDE. Pogamut 2 features several distinct tools supporting the process of agent prototyping and debugging: The list of running servers and agents (1) helps with a management of multiple agents. The agents' properties inspector (3) gives a quick access to variables common for all agents, e.g. position, velocity, orientation, health. The logs (5) display logged messages from communication with the GameBots, the platform and the agents' logic logs. The UT04 game server can be remotely administrated from (6).

As said above, the goal of the project Emotions in Pogamut 2 is the extension of the educational potential of the platform by the emotion aspect. This should be achieved by simple (easy to understand) and extensible emotion model along with provided scenarios that will cover educational goals defined in the introduction.

Below the prototype of emotion model will be described along with the presentation of two forms of scenarios currently considered. This will be followed by a preliminary analysis of design decisions.

The emotion model. The emotion model is inspired by OCC theory [8]. Emotions in the model are represented by their intensities, which range from 0 to 1. The set of simulated emotions is limited compared to OCC theory. The set includes anger, fear, joy, sorrow and surprise. Surprise is an extension to OCC – it is modeled by variable expectedness. Other emotion affecting variables used in the model include desirability, praiseworthiness, appealingness and intensity (intensity supplements most of the intensity affecting variables of the OCC model).

Each emotion in the model has a valence. The valence is a variable with three states – positive, negative or neutral – and is used for determining the mood of the agent. The current mood is counted as the difference between averages of all emotions with positive valence and all emotions with negative valence (emotions with neutral valence are not affecting the mood).

The model is equipped with a limited memory of encountered agents and situations. The averages of changes of emotions intensities caused by the agent or situation are stored in the memory. This information then influence event appraisal. The model also features a mechanism of habituation that is used especially for lowering the intensity of changes of emotions caused by situations that occur repeatedly. Memory and habituation mechanism help the model to adapt to the environment.

The model works in three phases (Fig. 3). Firstly, the emotion-significant situations in the environment are recognized (the recognized situations can differ according to the scenario). The situations will be defined in Pogamut 2 platform. Each time a situation occurs in the environment, it will be sent to the emotion model as a data structure containing the necessary properties (who participates in the situation and what has actually happened). Secondly, the situation is appraised by a set of OCC variables. The value of these variables will depend on the properties of the situation and on the agent's current state and goals. Thirdly, these variables are processed by the set of IF-THEN rules specified by the OCC theory, resulting in potential change of emotions intensities.

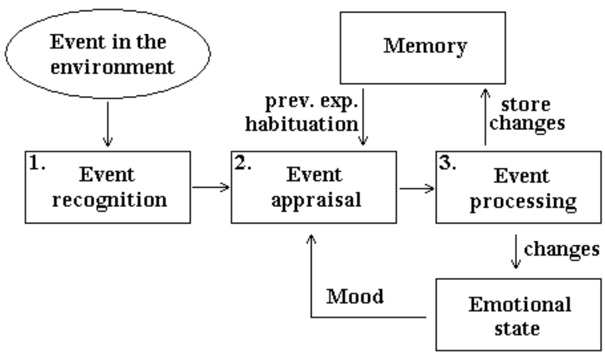


Fig. 3. The emotion model architecture. Description in the main text.

The resulting emotions intensities are influenced by the agent's current mood, his previous emotional experience with the situation, emotions attributed to other agents that participate in the situation, and the mechanism of habituation. Previous experience will influence emotions for example as follows: If some situation caused agent to be angry in the past, he will have higher tendency to get angry in the future experiencing the same situation.

The eyewitness agent scenario. This scenario will present one agent with emotions – the main character, a set of background actor-agents without emotions and one tutor without emotions. The main character will be eyewitnessing a set of predefined situations performed by actor-agents, which will be designed to trigger emotion response. Examples of such situations include other agents winning money in a lottery, a wedding, a bank robbery or even a car accident. The main character will comment these situations by a text module, telling a story about it. The changes of his emotions will affect the way he is commenting. This will be then explained by the tutor. Students will be able to interact with the scenario – they will be able to change the parameters of the emotion model and guide the main character through the environment.

In the scenario, the OCC variables of each situation will be shown along with resulting changes in the agent emotion intensities. The speech text module will be implemented using templates. Mapping of emotions to expressive behaviors in this scenario will be done one to one (one emotion for one type of expressive behavior).

This scenario can be used to explain emotional event appraisal for OCC based models. It also shows one of the roles of emotions for virtual agents – expressive behavior and emotions affecting the speech – together with associated issues (e.g. believability of expressive behavior against believability of emotion model).

The emotion agent scenario. This scenario will show that emotions can be used in agent decision making and demonstrate the advantages and pitfalls of this approach. The goal is to show properties of behaviors of agents with emotions and contrast them with behaviors without emotions. The emotions will be used as a control mechanism – agents' behaviors will be triggered exclusively by them.

The scenario will be aimed on agent survival (for this purpose the environment of UT04 is ideal as it is an action game which contains plenty of mechanisms and rules aimed on survivability). The agents will play regular game of UT04. They will try to conquer the opponent as often as possible trying to survive themselves.

Each emotion will be associated to one type of behavior (e.g. fear – avoidance behavior). The behavior with the emotion of highest intensity will be triggered. This will require a filter that will prevent rapid periodic switching of behaviors.

We remark that the aim of the scenario is not to show that the agent with emotions will score better than the agent without emotions. The scenario will show that the emotions can be used as condensed information about the environment and that they can take part in agent's decision making.

3 Analysis and Discussion

Platform Pogamut 2 and UT04 have been chosen because of two reasons. Firstly, as Pogamut 2 is our previous project, we have a good experience working with the platform and good idea how to extend it. Secondly, and more importantly, advantage of UT04 is that it can be easily extended because of its scripting language (UnrealScript), which is used for programming game mechanics. Epic Games (creator of UT) is still working on Unreal Tournament game series (UT 2007 was recently published) providing the game with better graphics engine, but still using UnrealScript language. This enables the migration of our projects to new versions of UT (taking advantages of facial expressions support and better gesture support).

The motivation to base emotion model on OCC is that the OCC model is often used as a theoretical background or it is even implemented in projects involving virtual agents, e.g. [10, 11, 12]. Its limitations are well known [13] and enhancements of the model are still being discussed [14]. Overall we believe that OCC theory is a good starting point in education of problematics concerning emotion modeling.

The removal of some emotions in the implemented model is justified by two facts. Firstly, as stated by Ortony in [14], the OCC theory by itself contains unnecessarily high number of emotion categories (concerning modeling emotions for virtual agents). Secondly, as the complexity of virtual environment is limited (comparing it to the real world), it is possible to limit the emotion model (e.g. the number of modeled emotions). Moreover the limited set of modeled emotions will enable easy mapping of emotions to expressive behavior (one to one). It will be possible to extend the model by the rest of OCC emotions – if needed (by adding the necessary variables and by extending event appraisal to support these variables).

There are several open issues. Firstly, there is the question whether we should use BDI (Belief Desire Intention) approach for controlling the agent or whether simple IF-THEN rules will be sufficient? This decision will be affected by the complexity of the emotion model. The model should be simple and easily understandable and at the same time it should show variety of emotion mechanisms. We plan to use both approaches (IF-THEN rules and BDI) and evaluate which of them will suite the problem better.

Secondly, we believe that the major advantage of interactive scenarios is that they provide feedback on ideas of students. The question is the amount of the interactivity needed for the efficient education. We will produce several scenarios with different amount of interactivity. Then we will be able to see which of them will be prefered by the students and we will have grounds for more thorough evaluation.

Thirdly, there is a question of the manifestation of the emotions. This is a complex problem by itself. For the project we need a set of clear-cut emotion manifestations. UT04 provides number of gestures and animations and can be extended by new ones. Facial expressions support is somehow limited in UT04, so we have decided not to include it in the project. For the project it may be needed to extend UT04 by the module supporting emotional expressivity (the work on this module already began). There is also the possibility of migration to Unreal Tournament 2007, which provides better gesture and facial expressions support. In future, we will also consider a possibility of defining gestures through BML [15], which is a markup language that provides standardized way of gesture specification.

4 Conclusion

The paper presented an ongoing project aimed on the education in the field of emotion modeling using virtual environments. The project will extend educational platform Pogamut 2, which uses virtual environment of the action game Unreal Tournament 2004. The goal of the project is to complement theoretical education in the field of virtual agents and emotion modeling by interactive scenarios demonstrating the problematics. An eyewitness and emotion agent scenarios were proposed to cover the following topics: roles of emotions for virtual agents (emotions and expression, emotions as information, emotions and control) and introduction to OCC like models (general overview of the mechanisms used, event appraisal).

Future work will include the implementation of the emotion model along with creating the scenarios in UT04. Then the project will be validated in the ongoing course at Charles University in Prague concerned with virtual agents.

Acknowledgments. This work was partially supported by the Program "Information Society" under project 1ET100300517, by the Ministry of Education of the Czech Republic (Res. Project MSM0021620838), and by student grant GA UK No. 1053/2007/A-INF/MFF. We would like to thank Klára Pešková and Rudolf Kadlec for their contributions to this paper.

References

1. Damasio, A.: Descartes' Error, G.P. Putnam's Sons, New York. (1994)
2. Aylett, R.S., Louchart, S., Dias, J., Paiva, A., & Vala, M.: Fearnot! - an experiment in emergent narrative. In: Proc. IVA 2005, pp. 305--316. Springer Verlag LNAI 3661. (2005)
3. Bída M., Brom C.: Emotion virtual agents. In: Proceedings of Kognice 2007. Olomouc, Czech Republic. In print. (in czech) (2007)
4. Bevacqua, E., Mancini, M., Niewiadomski, R., & Pelachaud, C.: An expressive eca showing complex emotions. In: Proceedings of AISB'07. Newcastle University, Newcastle upon Tyne, UK. (2007)
5. Bevacqua, E., Heylen, D., Pelachaud, C., & Tellier, M.: Facial feedback signals for ecas. In: Proceedings of AISB'07. Newcastle University, Newcastle upon Tyne, UK. (2007)
6. Kadlec, R., Gemrot, J., Burkert, O., Bída, M., Havlíček, Brom, C.: POGAMUT 2 - A platform for fast development of virtual agents behaviour. In: Proceedings of CGAMES 07, La Rochelle, France. (2007) http://artemis.ms.mff.cuni.cz/pogamut/ [29.7.2008]
7. Epic Games: Unreal Tournament 2004. http://www.unrealtournament.com [29.7.2008]
8. Ortony, A., Clore, G.L., Collins, A.: The cognitive structure of emotions. Cambridge University Press, Cambridge, UK. (1988)
9. Adobbati, R., Marshall, A. N., Scholer, A., and Tejada, S.: Gamebots: A 3d virtual world test-bed for multi-agent research. In: Proceedings of the 2nd Int. Workshop on Infrastructure for Agents, MAS, and Scalable MAS, Montreal, Canada. (2001) http://www.planetunreal.com/gamebots [29. 7. 2008]
10. El-Nasr, M.S., Yen J., Ioerger, T.: FLAME - A Fuzzy Logic Adaptive Model of Emotions. In: Automous Agents and Multi-agent Systems, Volume 3, pp. 219--257. (2000)
11. Gratch, J., Marsella, S.: A Domain-independent framework for modeling emotion. In: Journal of Cognitive Systems Research, Volume 5, Issue 4, pp. 269--306. (2004)
12. Romano, D.M., Sheppard, G., Hall, J., Miller, A., Ma, Z.: BASIC: A Believable, Adaptable Socially Intelligent Character for Social Presence. In: PRESENCE 2005, The 8th Annual International Workshop on Presence, pp. 21--22. University College London, London, UK. (2005)
13. Bartneck, C.: Integrating the OCC Model of Emotions in Embodied Characters. In: Workshop on Virtual Conversational Characters: Applications, Methods, and Research Challenges, Melbourne. (2002)
14. Ortony, A.: Emotions in Humans and Artifacts, chapter On making believable emotional agents believable, pp. 189--212. MIT Press. (2003)
15. Vilhjalmsson, H., Cantelmo, N., Cassell, J., Chafai, N., Kipp, M., Kopp, S., Mancini, M., Marsella, S., Marshall, A., Pelachaud, C., Ruttkay, Z., Thorisson, K., van Welbergen, H., & van der Werf, R.: The behavior markup language: recent developments and challenges. In: Proceedings of IVA '07. Paris. (2007)

Demonstrations

presented at the workshop

Interpreting the Dynamics of Facial Expressions in Real Time Using Model-based Techniques

Christoph Mayer, Matthias Wimmer, Freek Stulp, Zahid Riaz, Anton Roth, Martin Eggers, Bernd Radig

Technische Universität München
Boltzmannstr. 3, 85748 Garching

1 Introduction

Recent progress in the field of Computer Vision allows intuitive interaction via gesture or facial expressions between humans and technical systems. Our demonstration shows model-based recognition of facial expressions in real-time, see Figure 1. Yet, its complexity challenges the task of model fitting and we tackle this challenge with an algorithm that has been automatically learned from a large set of images. Note, that our system is not limited to facial expression estimation. Gaze direction, gender and age are also estimated.

2 Model-based Facial Expression Estimation

Models rely on a priori knowledge to reduce the large amount of image data to a small number of model parameters which facilitates the subsequent interpretation task.Cootes et al. introduced modeling based on variations in shapes and appearance [2]. Recent research considers modeling faces in 3D space [1, 9]. Model fitting is the challenge of determining model parameters that best describe the face within a given image and usually relies on three main components [8].

Fig. 1. The complex state-of-the-art Candide-3 face model consists of 116 landmarks and reflects the face shape by 79 shape parameters.

The face model contains a parameter vector $\mathbf{p}$ that represents its configurations. We integrate the deformable 3D wire frame Candide-3 face model [1]. *The localization algorithm* computes an initial estimate of the model parameters that is further refined by the subsequent fitting algorithm. Our system integrates the approach of [7].*The objective function* yields a comparable value that specifies how accurately a parameterized model matches an image. In contrast to traditional approaches that manually specify the objective function, we automatically learn it from a large set of training data.As a result, this approach yields more robust and accurate objective functions.*The fitting algorithm* searches for the minimum of the objective function. However, the focus of our work is on the objective function rather than on the fitting algorithm.

As described by Ekman [5] face structure indicates facial expressions. In a reverse conclusion we infer facial expressions currently visible from it. In addition, we take the motion of certain landmarks that are connected to the face structure within visible faces into account. This large amount of feature data provides a profound basis for the classification step, which therefore achieves great accuracy. A classifier infers the facial expression visible in the camera image.Our demonstrator recognizes the major human facial expressions.

3 Discussion and Conclusion

High quality objective functions that are learned from annotated example images ensure both an accurate and fast computation of the model parameters. Our demonstrator for facial expression estimation has been presented at several events with political audience and on TV. The drawback of our approach is that the data base from which the objective function is learned needs to cover all aspects of face properties.An online fitting demonstration is available[1].

References

1. J. Ahlberg. Candide-3 – an updated parameterized face. Technical Report LiTH-ISY-R-2326, Link 2001.
2. T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *ECCV 1998*, Springer-Verlag.
3. T. F. Cootes and C. J. Taylor. Active shape models – smart snakes. *BMVC 1992*, pages 266 – 275. Springer Verlag.
4. P. Ekman. Universals and cultural differences in facial expressions of emotion. *Nebraska Symposium on Motivation*, 1971
5. P. Ekman and W. Friesen. *The Facial Action Coding System: A Technique for The Measurement of Facial Movement*. Consulting Psychologists Press, 1978.
6. T. Kanade, J. F. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *Int. Conference on Automatic Face and Gesture Recognition*, 2000.
7. P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 2004.
8. M. Wimmer, F. Stulp, S. Pietzsch, and B. Radig. Learning local objective functions for robust face model fitting. *IEEE (PAMI)*, 30(8), 2008.
9. V. Blanz, K. Scherbaum, H. Seidel. Fitting a morphable model to 3d scans of faces. *ICCV07*.

[1] http://wwwradig.cs.tum.edu/people/wimmerm/demo