A Model for the Generation and Combination of Emotional Expressions

Aldo Paradiso
GMD - German National Research Center for Information Technology Institute (IPSI)
D-64293 Darmstadt, Germany
+49.6151.869843
paradiso@darmstadt.gmd.de

Marcello L'Abbate
GMD - German National Research Center for Information Technology Institute (IPSI)
D-64293 Darmstadt, Germany
+49.6151.8694945
labbate@darmstadt.gmd.de

ABSTRACT
In this paper we introduce a model, called Expression Generator, which generates facial expressions related to pseudo emotions and behavior of chatterbots, which are characters used as front ends of dialogue systems employed in Web sites. The emotions are articulated as sequences of MPEG-4 compliant facial displays, and visualized in a browser-embedded player. We define a set of operators to combine facial displays and propose an algebraic model to control the stream of expressions.

Keywords
Multimodal interfaces; conversational interfaces; facial animation.

1. INTRODUCTION
Research on multimodal interfaces has become quite a tradition. One important instantiation of multimodal interfaces uses the metaphor of face-to-face conversation and has been applied to human-computer interface design for quite some time. Several authors have pointed out the major trends in such a broad field [14, 2]. In this paper we describe our approach in the development of an emotion-sensitive character embedded in a dialogue system being employed in the COGITO project. This EU-funded project aims at improving consumer-supplier relationships in future e-commerce interfaces by featuring agents, which can converse with users in written natural language ("chatterbots") [17]. Whereas contemporary graphical interfaces are usually „passive", the COGITO project aims at building a system that is not merely reactive to some user request, but pro-active and capable of engaging in a goal-directed conversation with the user. The naturalness of interaction, especially for casual users, is enhanced by suitable 2D and 3D animations of the agent expressing appropriate emotional reactions.

The concept of a chatterbot is central to an interactive and conversational system. It consists of a software system that attempts to simulate the conversation or "chatter" of a human being, often entertaining the user with some "smalltalk". Chatterbots such as "Eliza" and "Parry" are well-known early attempts at creating programs that could at least temporarily fool a real human being into thinking they were talking to another person. A chatterbot can be thought of as the spokesperson for an artificial intelligence system, usually consisting of a collection of dialogue management rules, which use different techniques for processing the user's input. These techniques may range from simple keywords based text parsing to more complex logic-oriented methodologies based on inference mechanisms. Applying a rule may have mainly two effects: first, the output text is determined, and presented to the user possibly in combination with a graphical visualization, e.g. a cartoon displaying the related emotion. Second, a rule may cause a changing of the chatterbot's internal state. In this way the rule base may be subdivided into different contexts, each of which is concerned with a particular topic. The management of sentences having different meanings depending on the subject of conversation may therefore become more flexible and efficient. Within the Internet, chatterbots are commonly used as interactive guides during a web site tour (see www.extempo.com for on-line examples). While showing the different pages of the web site, the virtual agent tells the user about the functionality and the main controlling mechanisms of the currently displayed page.

The chatterbot that we employ is called eBrain. Its main purpose is the application within e-commerce web sites [7]. It may be considered a virtual employee of the companies using it. Besides the usual "smalltalk" capabilities, eBrain is mainly applied for a natural language guided tour of the web site contents and for the overall presentation of the company. This comprises of course also the choice of products the company is offering. In addition to the already mentioned subdivision of the rule base in contexts, a set of global variables is also managed. Their values are updated and made available during the entire dialogue, allowing a first restricted personalization. For instance, variables containing the name and gender of the user can be accessed by any rule and used for composing the output sentences.

Having introduced the concept and role of chatterbots, in the next section we will discuss on how to generate emotions and behavior related to an on-going dialogue with a chatterbot.

2. GENERATION OF SYNTHETIC EMOTIONS
Automatic generation of emotions belongs to a multidisciplinary field including perceptive psychology, cognitive sciences, knowledge systems, and facial and body animation. Dialogue systems featuring autonomous characters have been already designed and implemented in the past ten years [1, 2]. B. Hayes-Roth describes characters provided with a personality and a role, and immersed in a recognizable context [8].

Her approach is noteworthy also because of the fact that an authoring system has been developed, where non-professionals can easily define what she calls the "personality, role, and application context". Though this approach has its value from the point of view of personalization of characters, little efforts have been made to create convincing facial animations of the characters. Also, there is no speech synchronization.

Table 1 - Mapping of expressions in MPEG-4 (i.e. values of FAP 2)

FAP value	expr. name	textual description
1	joy	The eyebrows are relaxed. The mouth is open and the mouth corners pulled back toward the ears.
2	sadness	The inner eyebrows are bent upward. The eyes are slightly closed. The mouth is relaxed.
3	anger	The inner eyebrows are pulled downward and together. The eyes are wide open. The lips are pressed against each other or opened to expose the teeth.
4	fear	The eyebrows are raised and pulled together. The inner eyebrows are bent upward. The eyes are tense and alert.
5	disgust	The eyebrows and eyelids are relaxed. The upper lip is raised and curled, often asymmetrically.
6	surprise	The eyebrows are raised. The upper eyelids are wide open, the lower relaxed. The jaw is opened.

Once emotions are generated, there is the need to properly correlate them to facial expressions. Descriptions of emotions related to facial expressions have been carried out since 1952, where Schlosberg recognized six major emotions (i.e. happiness, surprise, fear, sadness, anger, disgust), acknowledged from pictures of faces, and put them in a 2D space. Each emotion occupied a defined region, and put in relationship with all other emotions in a circular way [12]. This interesting work has been the ground for implementations of systems for the generation of facial expressions related to emotions as well as new approaches for HCI in GUIs [4].

The MPEG-4 facial animation system addresses the integration of visemes and facial expressions [9, 10]. MPEG-4 is an object-based multimedia compression standard, which allows for encoding of different audio-visual objects of a scene independently. In particular, MPEG-4 enables the integration of facial animation with multimedia communications and presentations and allows facial animation over low-bandwidth communication channels.

In this approach, each facial expression is coded as an ordered sequence of 68 integer numbers, called Facial Animation Parameters (FAPs). Each of the parameters 3-68 acts on points defined on a synthetic face (feature points), and each point governs the deformation of its surrounding area, resembling muscle movements. This approach is an evolution of the FACS coding system developed by Elkman and colleagues in 1972 [5]. The remaining FAPs 1 and 2 refer to viseme and expression coding respectively. These are high-level parameters. Parameter 1 can code 14 visemes, while parameter 2 can code the six emotions mentioned above (see Table 1), which are universally accepted by the scientific community to represent a wide and culturally independent set of facial expressions [6].

In this work we propose a framework with which we a) describe a set of emotional motivators that come out of a conversation generated by a Dialogue System and propose a method to detect and quantify them; b) integrate the representation and animation of facial displays and possibly speech postures in the same stream, basing our work on the MPEG-4 standard and therefore propose an extension of it.

The communication channel considered in this paper, however, is unidirectional, that is, we only consider the generated facial displays and speech shown on the synthetic face of a character. We do not take in consideration any input from the user, except the text sent via a keyboard. Moreover, we concentrate on the head of the model, at least in the implementation part, even if we acknowledge that body gestures can play important roles as communicative acts.

3. EMOTION GENERATION IN THE COGITO PROJECT

The eBrain chatterbot introduced in the first section is a core component within the Cogito project [17], which provides a suitable application framework for demonstrating the usefulness of an expressive chatterbot. Generally, the Cogito project aims at improving the intelligence and efficacy of a virtual agent employed as the main interface to a commercial web site. In order to achieve this, the chatterbot is expanded with a combination of several technologies, including personalization mechanisms based on machine learning techniques [13], an automatic expression generator and a 3D synthetic expressive cartoon-like face able to show emotions and, later on, speech capabilities. By exploiting the framework defined in the Cogito project, we identified a set of parameters that may contribute to the generation of the synthetic emotion. We call these parameters motivators and they are the followings:

Keywords

Keywords are particular words or expressions appearing during the dialogue can be directly related to a given emotion. For instance a laughing face may accompany the word "happy". In our mapping, we are able to associate keywords to: a) emotions, which are not necessarily restricted to the six major ones (i.e. joy, sadness, anger, fear, disgust, surprise); and to b) comments, which occur when the speaker wants to emphasize linguistic items with head nods, and/or with eyebrows raises (ex: in "I really mean that!" nodding and raising eyebrows when saying "really"). Emotions and comments need both to be identified in the set of rules defined for an eBrain instance. For that, special tags will be used. The tags will be analyzed and filtered by the dialogue system, and a proper message will be sent out, synchronized with the textual output.

Context

The virtual agent's knowledge base is divided into contexts, each of them related to a particular topic of discourse. For instance talking about a natural disaster (like an earthquake) could imply excluding a happy expression. The context, as outlined at the end of section 1, is a set of rules, related to each other by a common topic. The whole set of contexts is arranged in a hierarchical way, e.g. some contexts can be reached only by first considering precedent ones in the hierarchy. For instance if you consider a context dealing with music issues, it will be most likely to have other sub contexts dealing with a specific kind of music such as rock or classic music. A context change will be applied whenever a particular keyword is recognized within the user's input. By still considering the previous example, if the user mentions the term "rock" while generally talking about music, the subordinated context about rock music will be invoked and the contained rules processed. If the user's input is not recognized within the current context, e.g. if there is no subordinated context dealing with the mentioned topic, the system will apply the so called "fallback" procedure, which means going up into the hierarchy through the superordinated contexts and checking whether there is a context that could be applied (see also [19]). In our model we track the current context. The analysis of the current context as well as of the user's input text (and therefore the new context) will produce as a result an index indicating the degree of coherence among the two of them. The value of such an index depends on whether the new context is subordinated or reached by fallback and how distant the new context is from the precedent one. For example, if the conversation is about books, and the user suddenly determines a context change by fallback asking about the current weather, this will produce an output value equivalent to a low coherence. This index may be implemented as a real number ranging from 0 to 1, 0 corresponding to no coherence at all, and 1 corresponding to maximum coherence, which is supposed to be the default value. The next step is to map the coherence index to some facial display. For simplicity we decided to relate the coherence to an expression signaling "doubt". The relationship is inversely proportional, so that a low coherence will be mapped to a high degree of doubtfulness, and vice versa.

Is it also possible to define an inherent emotion, or at least an emotional attitude, to be associated with each context. If we are talking about a funeral of someone who was close to us, the inherent emotion associated will most probably be sadness. This piece of information can be represented as an index of inhibition, which impedes the character to manifest too funny expressions.

User Profile

It is generally known that user profiles can be used to improve HCI [3, 14]. User profiles usually contain information about the user's age, gender, language spoken, geographical location, job, education, hobbies, etc. All of these attributes may affect the production (and visualization) of the character's facial displays. In our model we will take into consideration the age of the user only. We argue that there is an inversely proportional relationship between the user's age and the intensity of expressions (performed by our character). The expressiveness of the character will be stronger the younger a user is. For instance if the user is a child, the agent may tend to have more laughing and friendly expressions. Therefore we associate an inhibitor/amplifier index to the user profile affecting the facial display of the chatterbot.

Biological needs

Humans have normally visible biological needs, which also play a role during conversations. Typically they include wetting the lips or blinking with the eyelids and moving the head to relax muscles. Reproducing behavioral aspects of biological needs greatly improves the naturalness of a character. We will synthesize them by means of complete or partial FAP streams.

4. REPRESENTING EMOTIONS AND BEHAVIOR WITH MPEG-4 FAP STREAMS

In the previous section we have identified a set of motivators that can give a contribution on defining pseudo emotions. The next step to be performed is finding an effective way to describe and combine them seamlessly in order to obtain sequences of facial displays played onto the character, according to the nature of the conversation.

We base our description on the MPEG-4 standard. We can then distinguish two levels of FAP: a low-level one, where each FAP defines the position of feature points onto the face (or the rotation of some facial objects like eyes and neck), and a high-level one (i.e. FAPs 1 and 2), where each FAP describes the whole face. Since the values of FAPs are simple integers, it is then necessary for high-level FAPs to map these numbers to a set of low-level FAPs; in other words, it is necessary to interpret them. The interpretation task is not part of the standard, but left to the decoder system. Therefore, different decoders may use different interpretations, and produce expressions that can look different for the same high-level FAP values. Though, the MPEG-4 standard provides syntax to define intensities of expressions as well as a way to blend them (see Table 2). In particular, the expression parameter, composed of a set of values as shown in Figure x, allows blending together two expressions from a standard set. The expressions to be blended are defined with integer values *expression_select1* and *expression_select2*, while their intensities are defined with *expression_intensity1* and *–2*. The formula defined in the standard for blending the two expressions is the following:

Table 2 – Syntax of expressions in MPEG-4

expression () {	**Range**
expression_select1	0-6
expression_intensity1	0-63
expression_select2	0-6
expression_intensity2	0-63
…	…
}	

final expression = expression1 * (expression_intensity1 / 63)+
expression2 * (expression_intensity2 / 63)

The resulting value (final expression) will presumably be a set of FAPs that represent an expression interpolated between the given two. In this approach the concept of excitation (or intensity) of an expression is mixed with the one of combination of two expressions.

The approach adopted in [18] is based on MPEG-4, but it describes a different algorithm to blend two animations. Each FAP is provided with an intensity value over time as well as a weight. When two FAPs occur at the same time, a weighted FAP value is calculated, using these parameters, so that discontinuities in the animations are avoided. The concepts of intensity and weight introduced in this paper are similar to the ones we will introduce in section 5.

Nonetheless, according to the discussions taking place in MPEG-4 working groups, "using FAP 1 and FAP 2 together with low-level FAPs 3-68 that affect the same areas as FAP 1 and 2, may result in unexpected visual representations of the face" [16]. Therefore, it is recommended to use only those low-level FAPs that will not interfere with FAP 1 and 2. We will consider these recommendations, for the following reasons: A) FAP 1 and 2 seem not to be powerful enough to be used conjointly: what happens if we want a face talk and smiling at the same time? B) It is not possible to directly map the motivators (described in section 3) to values of FAP 2.

Therefore, we consider only the low-level FAPs 3-68 as suitable for the description of our facial displays.

In Figure 1 we can see a schema outlining the different layers involved in an emotion generation process. We first need to

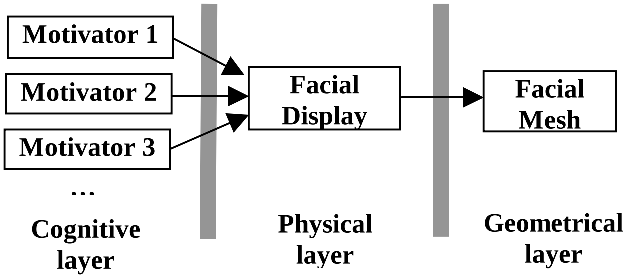

Figure 1 - A layered model of emotion generation

identify a set of motivators, then combine them into a facial display, and finally apply the result to the facial model. In the next section we will explain how we map the motivators to facial displays and how to merge them.

5. MAPPING MOTIVATORS TO FACIAL DISPLAYS

We are now in the position of making an association between dialogue elements and character's emotions/behavior. We need to relate the emotions and behavior to visible facial displays. In other words, we need to know what facial expressions we need to produce, when, and for how long. The distinction between emotions and facial displays, though apparently clear, is somewhat hidden in the definition of expression itself. For example, in Table 1 the emotions are described in terms of facial expressions, even though a) this description is just one possible interpretation of the emotion and b) facial expressions are only one of the possible vehicles of expression, together with speech prosody and intonation, gestures, and others.

Table 3 - Mapping of behavior to facial displays

Behavior/emotion	**Facial Displays**
Showing interest	1 - Raise eyebrows 2 - Rolling the head slightly
Being bored	1 - Looking ahead 2 - Closing eyelids longer
Do not understand what the user says	1 - Inner eyebrows pulled downward and together 2 - turning the head to show one ear meaning "What?"
I'm thinking	1 - Inner eyebrows pulled downward and together, and looking ahead, or downwards.
joy, sadness, anger, fear, disgust surprise	See Table 1

Cassell [2] makes a distinction between conversational functions and conversational behaviors. The former being rules we follow when performing conversations and related behavior, the latter being all the acts we physically perform to manifest the functions. For example, raising the eyebrows may mean emphasizing a word (i.e. a concept) in a sentence. We will use an approach similar to Cassell's for making a mapping between behavior and facial displays (Table 3). This mapping incorporates the expression descriptions of Table 1.

Table 3 is not complete, and it may be extended and refined. Currently, we are able to deal with emotions and other behavior in terms of FAP streams, which can be defined, stored, retrieved and executed at the client side.

We can now face the issue of composing facial displays. In our model a motivator may affect the production of a facial display in different ways. First of all, it may be associated directly to a facial display, that is, mapped directly to a FAP stream.

For instance, we can associate the facial display "smile" to a keyword "happy". The association may also be partial, that is, only a subset of the FAP stream will be associated. For instance, in the sentence "I really like it" the eyebrows will be raised in correspondence to the word "really". A third association concerns the production of weights to a motivator. A weight is a rational scalar value, typically between 0 and 2, which will be multiplied to a FAP stream (a vector) in order to enhance or inhibit its intensity or visual effect. For instance, we may apply a weight of 0.5 to a smiling face in order to inhibit the smile. Weights may also be applied to a subset of FAP stream. Table 4 shows a list of motivators and their effects to facial displays. Facial displays will be revealed according to a certain dialogue state, and triggered by one or more motivators. The final expression and its duration, however, depend on a combination of different displays, which

Table 4 - Motivators and their effects to facial displays

		Facial Display			
		Complete FAP streams	Partial FAP streams	Whole Weights	Partial Weights
Keywords	Emotions	✓		✓	
	Comments		✓		✓
Context		✓	✓	✓	✓
User Profile (age)				✓	✓
Biological Needs		✓	✓		

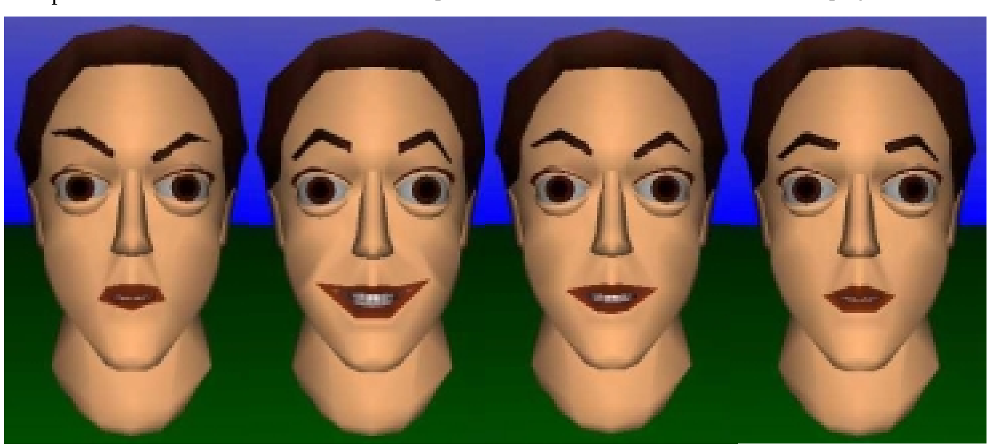

Figure 2 - Merging expressions. In order we have: anger (a), smile (b), their weighted sum (c), and the neutral expression (d).

may be produced at the same time. For example, the effect of biological needs and user profile lasts usually for the whole session, and this must be combined with the ongoing expressions triggered by keywords and/or current context.

For this motivation it is necessary to define a criterion to combine the effect of different motivators. We need to define operators to combine FAP streams.

If we have two facial displays $E=(e_3, \dots, e_{68})$ and $F=(f_3, \dots, f_{68})$, e_i, $f_i \in Z$ (natural numbers, not considering FAP 1 and 2, for the motivations explained in section 4), we define the sum operator $+_v$ as the following facial display:

$$+_v(E, F) = (e_i \cdot v + f_i \cdot (1-v)) \qquad i=3, \dots, 68 \text{ and } v \in [0,1]$$

The result is a FAP stream, which may be combined with other facial displays using the same operator, and so on. $+_v$ does not hold the associative property, nor the commutative one. If $v=0.5$ this operator will produce a facial display being the mean of the first two ones. Figure 2 shows an example of use of the sum operator over two expressions (anger and smile) applied onto the same face, with $v=0.5$.

As we stated before, a weight $w \in \Re$ is a scalar operator so that, if we have a facial display $E=(e_3, \dots, e_{68})$, then:

$$E \cdot w = w \cdot E = (w \cdot e_1, w \cdot e_2, \dots, w \cdot e_{68}) \qquad (1)$$

In Figure 3 a weight of 2.0 has been applied to a smiling expression.

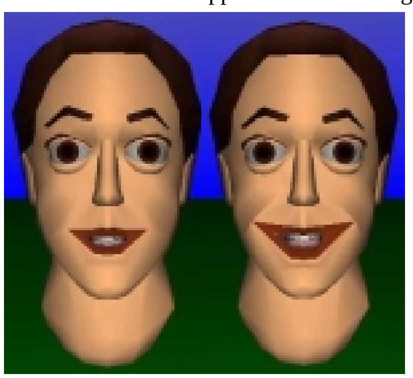

Figure 3 - Emphasizing a smile

It may happen that a facial display must occur integrally (and not weighted with other displays) even if there are concurrent factors producing different displays. For example, we need to wipe out our eyes regularly, even if we are screaming or we want to smile. Also, we can observe that facial displays may overlap without interfering each other because they affect different facial regions or movements. I can turn my head while smiling, and while closing the eyelids. To model these phenomena we have to first characterize facial displays in terms of priority and affected facial regions. Let us introduce mask vectors $M=(m_3, \dots, m_{68})$ where $m_i=0$ or 1. Mask vectors will serve us to identify partial regions of the complete FAP vector without loosing generality, i.e. we deal with vectors having the same size in all our computations. If we have a facial display $E=(e_3, \dots, e_{68})$ the partial region identified by M is $E_M = M^T \cdot E$ where M^T is the transpose of M. Then, if we consider two facial displays E, F with masks respectively M_1 and M_2 and priorities p_1 and p_2, $p_1 \neq p_2$, we introduce the overlapping operator O so defined:

$$O(E_{M1,p1}, F_{M2,p2}) = (g_i) = \begin{cases} g_i=e_i+f_i & \text{if } e_i \text{ or } f_i = 0 \\ g_i=e_i & \text{if } p1>p2;\ g_i=f_i \text{ otherwise} \end{cases} \qquad (2)$$

where $i=3,\dots,68$.

By adopting this operator we are able to combine partial facial displays producing their whole effect in selected areas. Timing information will be considered as well, since each of the motivators produces a facial display lasting for a given interval. We need therefore to put each of them as a function of the time t.

Let us now put together all the elements concurring at forming a facial display. At the end we will formulate an algebraic expression.

First of all we must consider all the motivators that directly produce whole facial displays, i.e. keywords (emotions), context, and biological needs. If these motivators occur together, they must overlap in some way. If we use the symbol "$\underline{o}$" as the overlapping operator defined in *(2)* and $Fd(t)$ as the resulting facial display, we may write:

$$\mathbf{Fd(t) = [K(t) +_v C(t)]\ \underline{o}\ Bn(t)}$$

where K = Keywords, Bn = Biological needs; C = Context.

However, the value of $Fd()$ is amplified or inhibited by the other motivators, namely: User Profile (age). Keywords and context may also play such a role, but at the moment, for simplicity, let us ignore it. We must also consider that the user profile affects the whole session of conversation. In this model the user profile is the only motivator affecting the value of the weight w previously defined in *(1)*. Therefore we have $w = f(Up)$, where Up = User Profile. According to this new parameter we have:

$$\mathbf{Fd(t) = f(Up) \cdot [K(t) +_v C(t)]\ \underline{o}\ Bn(t)}$$

The nature of the function $f(Up)$ may be further discussed. In this model we only consider the user age. We argue, as a first trial, that the age is inversely proportional to the intensity of the expression, as stated in section 3.

Note that since the biological needs do not depend on conversational or emotional situations the motivator Bn is not affected by Up. Also, since the user profile effect remains constant during the whole session, we did not put any dependency on the time t. Of course, we need to specify constraints that impede to reach meaningless values of the equation. For example, $f(Up)$ should be $\neq 0$.

6. SYSTEM ARCHITECTURE AND PROTOTYPE

Figure 4 outlines the architecture of the Expression Generator (in the dotted box).

The **Motivator Extractor** extracts the motivators, i.e., age from User Profiles, Context from eBrain log files, keywords and tagged text from the eBrain server. It then passes these parameters to the **Multiplexer** module. Such a component will produce MPEG-4' FAP streams according to the formulas defined in the previous

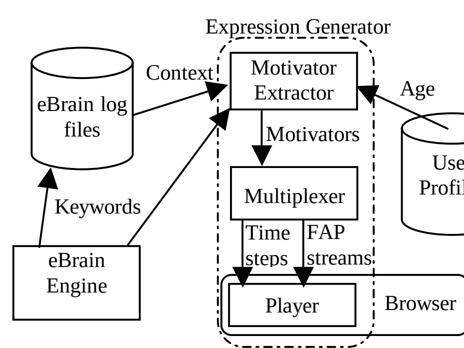

Figure 4 - System architecture of the Expression Generator

section, together with timing information, needed to synchronize the sequences of streams.

The **Player** module will then produce and run animations. The player is embedded in a HTML browser and based upon the Tinky system, a tool for the generation and animation of expressions on a 3-D cartoon-like VRML model [11]. In our approach, an animation is a sequence of expressions, together with a set of time steps occurring between them. We exploit the animation technique called keyframe animation, where a position or rotation is specified for only a few, key fractional times. The time steps defined in our animation are transformed in a set of key fractional times, and the FAPs streams (the expressions) are used to calculate a set of mesh values, transformed in key values. These values are considered as rough sketch of the animation, and the VRML interpolator nodes automatically fill in the values between those specified as needed. Thus, using keyframe animation, an animation description specifies only a few positions and rotations, instead of a large number of them.

We have built a prototype of the expression generator, which is currently being integrated in the eBrain client system.

7. FUTURE WORK

The system will be improved by adding several features, like:

a) Speech capabilities: the integration of speech will complete the expressiveness of the character. We plan to integrate a TTS system, synchronize the speech with labial movements of the character, and merge the visemes (i.e. the visual counterpart of phonemes) with concurrent facial displays.
b) An improved Dialogue Model. We intend to refine the existing dialogue model implemented in the eBrain system by adopting the so called Conversational Roles (COR) model, developed by Sitter & Stein, [15] derived from speech act theory.

Other improvements consist in refining our model by considering other information derived from user profiles, like gender, background, education, and possibly mood. We will also perform a set of experiments to test our actual model and use it with a sample of users.

8. REFERENCES

[1] André, E.; Rist, T.; Müller, J.; WebPersona: A Life-Like Presentation Agent for the World-Wide Web. In Knowledge-based Systems, 11(1):25-36, 1998.

[2] Cassell, J.; Sullivan, J.; Prevost, S.; Churchill E. Embodied Conversational Agents. The MIT Press, Cambridge, Massachusetts, 2000. ISBN 0-262-03278-3.

[3] Costabile, M. F.; Esposito, F.; Semeraro, G.; Fanizzi, N. An Adaptive Visual Environment for Digital Libraries. International Journal on Digital Libraries, Vol 2, N. 2+3,1999, (124-143).

[4] Eliens, A. Avatars in RIF. Proceedings of Face 2 Face Symposium, March 24, 2000, Delft, The Netherlands.

[5] Elkman, P.; and Friesen. W. V. Manual for the Facial Action Coding System. Consulting Psychologists Press, Inc., Palo Alto, CA, 1978.

[6] Elkman, P. The argument and evidence about universals in facial expressions of emotion. In H. Wagner and A. Monstead, editors, Handbook of Social Psychophysiology, pages 143-146. John Wiley, Chichester, 1989.

[7] Guttman, R.; Moukas, A.; and Maes, P. Agent-mediated Electronic Commerce: A Survey. Knowledge Engineering Review, June 1998.

[8] Hayes-Roth, B. Staffing the Web with Interactive Characters. In Communications of the ACM, 1999 N. 42

[9] ISO/IEC IS 14496-2 Visual, 1999.

[10] ISO/IEC IS 14496-1 Systems, 1999

[11] Paradiso, A.; Nack, F.; Fries G.; Schuhmacher, K. The Design of Expressive Cartoons for the Web – Tinky. Proceedings of ICMCS Conference, June 7-11 1999, Florence (Italy).

[12] Schlosberg, H. The Description of Facial Expressions in Terms of Two Dimensions. In Journal of Experimental Psychology, Vol. 44, No. 4, October, 1952.

[13] Semeraro, G., Ferilli S., Fanizzi N. & Abbattista F., Learning Interaction Models in a Digital Library Service. In M. Bauer (Ed.), UM2001 User Modeling: Proceedings of the Eighth International Conference, Lecture Notes in Computer Science, Springer:Berlin, 2001 (to appear).

[14] Shneidermann, B. 1998. Designing the User Interface. Addison Wesley. ISBN 0-201-69497-2.

[15] Sitter, S. and Stein, A. Modeling information-seeking dialogues: The conversational roles (COR) model. RIS: Review of Information Science (online journal) 1 (1).

[16] Tekalp M., and Ostermann J. Face and 2-D Mesh Animation in MPEG-4. Available from http://www.cselt.it/leonardo/icjfiles/mpeg-4_si/8-SNHC_visual_paper/8-SNHC_visual_paper.htm

[17] Thiel, U.; Stein, A.; Semeraro, G.; Abbattista, F.; De Candia, L.; Fanizzi, N.; Candela, V.; Lops, P. & Valente, A., COGITO – E-Commerce with Guiding Agents Based on Personalized Interaction Tools. In Proceedings of the AICA Annual Conference, Taormina, Italy, September 27-30, 2000.

[18] Kshirsagar, S.; Escher, M.; Sannier, G.; M.-Thalmann, N.; Multimodal Animation System Based on the MPEG-4 Standard, Multimedia Modelling 99, Ottawa, Canada, October 4-6 1999, pp. 215-232.

[19] L'Abbate, M.; Thiel, U.; Chatterbots and Intelligent Information Search, in Proceedings of ECIR 2001, Darmstadt, Germany, April 4-6 2001, to be published.