Speaking with Emotions

E. Bevacqua

Department of Computer and System Science
University of Rome “La Sapienza”
elisabetta.bevacqua@libero.it

M. Mancini and C. Pelachaud

LINC - Paragraphe
IUT of Montreuil - University of Paris 8
m.mancini@iut.univ-paris8.fr
c.pelachaud@iut.univ-paris8.fr

Abstract

We aim at the realization of an Embodied Conversational Agent able to interact naturally and emotionally with user(s). In previous work [23], we have elaborated a model that computes the nonverbal behaviors associated to a given set of communicative functions. Specifying for a given emotion, its corresponding facial expression will not produce the sensation of expressivity. To do so, one needs to specify parameters such as intensity, tension, movement property. Moreover, emotion affects also lip shapes during speech. Simply adding the facial expression of emotion to the lip shape does not produce lip readable movement. In this paper we present a model that adds expressivity to the animation of an agent at the level of facial expression as well as of the lip shapes.

1. Introduction

With the development of 3D graphics, it is now possible to create Embodied Agents that can talk simulating that kind of communication that people know since they are born. Moreover, nonverbal communication is as important as verbal one. Facial expressions can provide a lot of information. In particular they are good window on our emotional state [9]. Emotions are a fundamental aspect of human life influencing how we think and behave and how we interact with others. Facial expressions do improve communication [13]; they can make clear what we are thinking, even without speaking. For example wrinkling our nose in front of something that we dislike communicates very clearly our disgust. Therefore, believable synthetic characters make the interaction between users and machine easier and more fulfilling, providing a more human-like communication. Experiments have shown that interface with facial displays reduces the mental gap between users and computer systems [25].

Most of the ECA systems developed so far have been concentrated in defining the appropriate nonverbal behavior linked to speech. But nonverbal behaviors are characterized not only by the signals of the expression itself but also by temporal parameters (e.g. time of appearance and disappearance of an expression) and by muscular activities quality (such as tense movement). In the aim of creating believable embodied conversational agent (ECA) able to serve as bridge in the communication between humans and the machine, ECA ought to be empowered with human-like qualities. In this paper we present a computational model of expressivity for facial expression and for the lip movements. The work presented in this paper is part of the Greta system that have been developed within the EU project MagiCster [1] . The system takes as input a text tagged with communicative functions information. The language of tags is called Affective Presentation Markup Language, APML [6]. APML is used as a script language to control the agent’s animation. To endow agent with expressivity quality we have extended APML.

In the next section we prevent an overview of the state of the art. We then present our extension of APML. We follow by describing the lip shape model.

2. State of the art

There is not a single way to approach the issue of emotions in ECAs. S. Kshirsagar and her colleagues have developed a 3D virtual characters with emotions and personality [8]. The agent can maintain a basic dialogue with users. Through a personality and emotion simulators the agent responds naturally to, both, the speech and the facial expressions of the user that are tracked in real time.

M. Seif El-Nasr, J. Yen and T. Ioerger [11] implemented a new computational model of emotions that uses a fuzzy-logic representation to map events and observations to emotional states. They based their work on the research on emotions that shows that memory and experience have a major influence on the emotional process.

A. Paiva and her research team approached the problem of modelling emotional states of synthetic agents

[1] IST project IST-1999-29078, partners: University of Edinburgh, Division of Informatics; DFKI, Intelligent User Interfaces Department; Swedish Institute of Computer Science; University of Bari, Dipartimento di Informatica; University of Rome, Dipartimento di Informatica e Sistemistica; AvartarME.

when implementing a computer game (*FantasyA*) where two opponents fight each other [14]. The battle is played in a virtual arena by two characters, one controlled by the user (through the doll *SenToy*), and the other by the system. They made their own decisions but are influenced by the emotional state induced by the user. So the agent's actions depends on its emotions, on its opponent's emotions, and on its personality. Carmen's Bright IDEAS [19] is an interactive drama where characters exhibit gestures based on their emotional states and personality traits. Through a feedback mechanism a gesture made by of a character may modulate its affective state. A model of coping behaviors has been developed by Marsella and Gratch [20]. The authors propose a model that embeds information such as the personality of the agent and its social role.

All those works aim to define models being able to decide what kind of emotion synthetic agents should "feel" in a particular situation.

Another area of the research on emotions in virtual characters is concerned with the expressivity of emotions.

M. Byun and N. I. Badler [3] proposed a method for facial animation synthesis varying preexistent expressions by setting a small number of high level parameters (defined based on the Laban notation [15] that drive low level facial animation parameters (FAPs from MPEG-4 standard [7]). The system is called FacEMOTE [3] and derives from EMOTE [4] which has been origanally developed for arm gestures and postures.

Very few computational models of lip shape take into account emotion. M. M. Cohen and D. W. Massaro developed a coarticulation model [5], based on Löfqvist's gestural theory of speech production [16]. To model lip shapes during emotional speech, they add the corresponding lip shapes of emotion to viseme definition [21].

Our work is somewhat related to facEMOTE but we do not modified directly the animation stream; rather we create a new FAP stream to animate our 3D agent from an input text where tags specifying communicative functions are embedded [24, 6]. Our approach differs from other ones as we have added qualifiers parameters to simulate expressivity in facial expressions as well as in lip movements. Moreover we drive our computational model of emotional lip shapes from real data [17].

3. Temporal characteristics of facial expressions

A facial expression is not only identified by a configuration of facial muscles but is also important how this configuration will be temporally activated.

For example we can consider, starting from an initial neutral state, the time (or, identically, the speed) needed for the facial muscles to contract and reach the final state corresponding to the final expression; we call it *onset* time. Similarly, the corresponding time needed for the muscles to change from their current state back to the rest position is called *offset* time.

So a single expression (in the sense of ”muscular configuration”) can assume different expressivity depending on the manner it appears (onset), the time it remains on the face (called the *apex* time) and finally the speed it disappears (offset). We have chosen this specification to represent the temporal characteristics of facial expressions [22]. One example of this representation is shown in Figure 1. The time is on the *x* axis while on the *y* axis there is the intensity of the expression; where 0 means ”muscles in rest position” and 1 represents ”muscles in final position”.

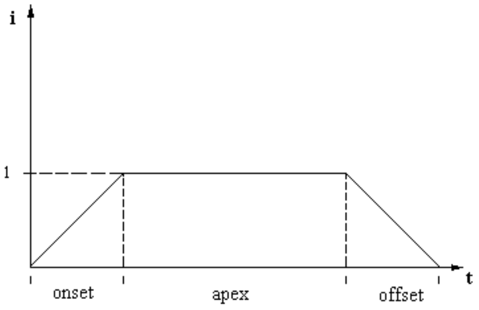


Figure 1: Temporal course of a facial expression.

Although these three parameters are extremely important in delivering expressivity of an expression, few studies assigning values to them exist; but vision systems could be used to extract them [12, 28].

It has been shown that, for example, expressions of sadness usually disappear slowly from the face. Thus, they have a long offset time [9]. On the other hand, expressions of joy appear very rapidly; this is characterized by a short onset time. Sometimes expressions may also differ in duration (apex time). Finally, fake expressions generally appear either too late or too early (e.g., polite smile versus real happiness smile); thqt is the value of delay varies.

4. Intensity of facial expressions

Facial movements corresponding to an expression can be produced with different intensity. An eyebrow can raise a little or a lot. A mild happiness will produce a small smile, while a great happiness will be shown by a large smile. Besides, P. Ekman found that if the emotion is felt very lightly, not every facial movement corresponding to the emotion will be visibly displayed; in the sense that changes expressed in the face may not be perceiv-

able [10]. For example, in the case of mild fear like apprehension, only slight expression around the mouth may be shown; while for extreme fear both areas, the muscles around the eyes and the mouth, are very tense [9].
Thus, the intensity of an emotion controls not only the amount of movements (strong or light) but also the appearance of some movements.

5. Giving expressivity to facial expressions

Until now we have been concentrating in elaborating computational models that define the most appropriate non verbal behaviors from an input text. The nonverbal behaviors as specified within APML correspond to frozen facial expressions; a facial expression is linked to an intensity; its temporal course is given by the XML span... Aiming at adding life characteristics to the agent, we have realized some modifications to the APML language in order to allow the Greta agent to communicate a wider variety of facial expressions of emotion as well as to allow for a more flexible definition of facial expressions and their corresponding parameters. Expressivity may be expressed differently depending on the considered modality: a face has different variables (timing variations, muscular intensity, as we have seen in sections 3 and 4) compared to gaze (length of mutual gaze, ratio of gaze aversion and looking at, ...). These modifications refer mainly to facial expressions timings as well as to their intensity.
An APML tag defines the meaning of a given communicative act [23]; the Greta engine looks up in a library of expression to instantiate this meaning into the corresponding facial expression. A facial expression has 3 temporal parameters as defined in Section 3: onset, offset and apex. In the previous version of the Greta engine, the value of the onset and of the offset were set as constants. An expression was set to start at the beginning of the tag and to finish at its end. That is the apex of an expression was set to be the total time length of a tag (computed as the duration of the speech embedded in the tag; this duration being provided by the speech synthesizer) minus the onset and offset times. APML provides a scheme to specify the mapping between meaning and signals for a given communicative act. We have extended APML to allow one to alter the expressivity of a communicative act. We have introduced a new set of 5 attributes that act both on expressions timings and intensities.
For each APML tag it is possible to specify one or more of the following attributes:

Delay : specifies the percentage of delay before an expression arises; it forces the Greta engine to delay the start of an expression for a certain time. This time is specified by a percentage of the total default animation time (that is the time of the speech embedded in the XML tag). If not specified, the default delay value is 0% (that is there is "no delay"); this corresponds to the previous version of APML.

Duration : specifies the total duration of an expression. It is specified as a percentage of the default expression duration. The Greta engine will set the duration of an expression (the apex of an expression) to last for this new value. The default value is 100% (that is "normal duration"); this corresponds to the previous version of APML.

Onset : specifies a value for the onset. This value is given as a number of animation frames that the engine have to use to render the "onset" phase of an expression. The default value is 0 and this tells to the engine to set the onset value to constant as defined in the previous version of APML and explained before.

Offset : specifies a value for the offset. This value is given as a number of animation frames that the engine have to use to render the "offset" phase of an expression. The default value is 0 and this tells to the engine to set the onset value to constant as defined in the previous version of APML and explained before.

Intensity : corresponds to a factor that multiplies the quantity of movement of each FAP involved in the facial expression. Until now the corresponding facial expression to a meaning for a given communicative act was explicitly defined. It was not possible to modify on the fly such a mapping. In order to have a facial expression with lower or greater intensity, one had to create a new entry in the library of expressions. This was quite cumbersome. To remedy this lack of flexibility, we introduce an intensity factor that can modify automatically any defined expression. Our facial model is MPEG-4 compliant [7]. An expression is defined by a set of FAPs (Facial Animation Parameter). The variation of their intensity corresponds to the modification of the value of each FAP. The default intensity value is 1 meaning that values of the FAPs defining an expression in the library of expressions are not changed.

Let us consider the following example:

```
<theme belief-relation="gen-spec"
    affect="anger">
some text </theme>
```

The timings of the expression as evaluated in the tag are given in the Figure 2.

Let us consider now the same example with the introduction of the new tags. For example we may have:

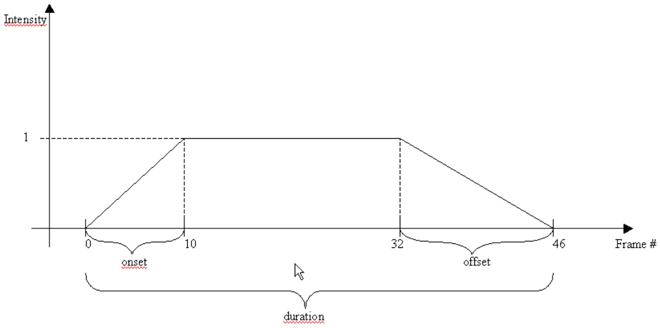


Figure 2: Temporal course of a facial expression.

```
<theme belief-relation="gen-spec"
    affect="anger" delay="40%"
    duration="40%" onset="4"
    offset="4" intensity="1.5" >
some text </theme>
```

The type of expression evaluated in the tag remains unchanged; what change are the temporal and intensity parameters of the expression. Figure 3 illustrates these changes. Figure 4 shows the resulting expressions as they can be seen on the agent's face.

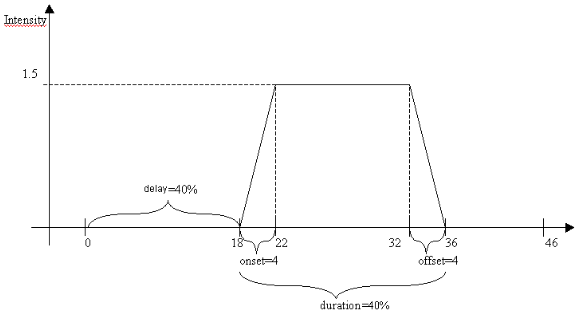
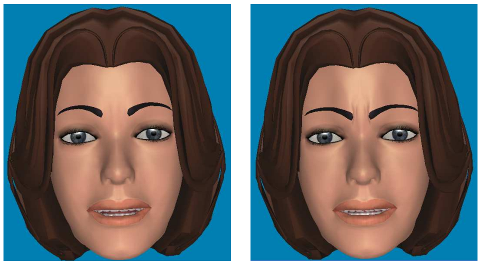


Figure 3: Temporal course of a facial expression when taking into account the new modifiers of APML.

Another example of how intensity can modify the appearance of an expression is given in Figure 5, in which an expression of joy is computed with an intensity of 1 in the first image and an intensity of 1.5 in the second.

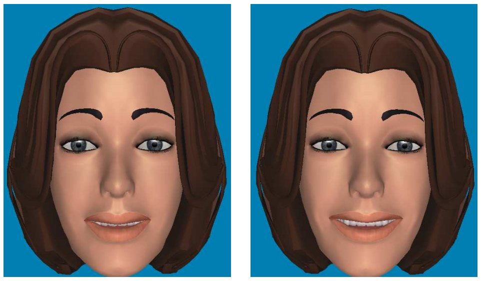
6. Computational lip model

Our lip model is based on captured data of triphones of the type $'VCV$ where the first vowel is stressed whereas the second is unstressed. The data has been collected with the optical-electronic system ELITE that applies passive markers on the speaker's face [17]. The data covers not only several vowels and consonants for the neutral expression but also different emotions, namely joy, anger, sadness, surprise, disgust and fear [17]. The original curves from the real speech data represent the variation over time of the 7 phonetically and phonologically relevant parameters that define lip shapes: upper lip height (ULH), lower lip height (LLH), lip width (LW), upper lip protrusion (UP), lower lip protrusion (LP), jaw (JAW) and lip corners (LC). On such curves we have selected the maximum or the minimum (target point) to characterize each viseme. To get a good representation of the characteristics of a vowel and of the breadth of its original curve, we choose two more points; one between the onset of the phoneme and its target and the other between its target and the offset of the phoneme. Instead, consonants are well represented considering just the target point. Since consonants, and at a slightest degree the vowels, are influenced by the context, we collect their targets from every ’VCV contexts (for instance, for the consonant /m/, the targets points are extracted from the contexts $/'ama/$, $/'eme/$, $/'imi/$, $/'omo/$ and $/'umu/$).



(a) intensity=1 (b) intensity=1.5

Figure 4: Anger expression: in figure (b) the frown is more intense, the lips are more tense and the teeth are clenched.



(a) intensity=1 (b) intensity=1.5

Figure 5: Joy expression: in figure (b) the cheeks are more raised, the lips are more widely open and the smile is larger.

Targets data of vowels and consonants have been stored in a database. Besides targets values, other information have been collected like the vowel or the consonant that defines the surrounding context, the duration of the phoneme and the time of the targets in this interval. A similar database for each of the six fundamental emotions and for the neutral expression have been created.

6.1. Lip shape algorithm

We have developed an algorithm that determines for each viseme the appropriate values of the 7 labial parameters by applying coarticulation and correlation rules in order to consider the vocalic and the consonantal contexts as well as muscular phenomena such as lip compression and lip stretching [2].

Our system works as follow. It takes as input a text (tagged with APML) an agent should say. The text file is decomposed by Festival [26] into a list of phonemes with their duration. The first step of our algorithm consists in defining fundamental values, called *target points*, for every parameter of each viseme associated with the vowels and consonants that appear in this phoneme list. A target point corresponds to the target position the lips would assume at the apex of the viseme (which may not always correspond to the apex of the phoneme production [17]). These targets have been collected analyzing the real data described above and are stored in a database, one for each emotion.

Afterwards, the algorithm modifies the targets according to the emotion in which the phonemes are uttered, to the coarticulation and correlation rules and to the speech rate.

Finally, the lip animation is computed on those targets.

	ULH	**LLH**	**JAW**	**LW**	**UP**	**LP**	**CR**
Neutral	0	0	0	0.2	0.1	0.1	0.2
Joy	1	1	1	0.8	0.9	0.9	0.8
Surprise	0	0	0	0	0	0	0
Fear	0	0	0	0	0	0	0
Anger	0	0	0	0	0	0	0
Disgust	0	0	0	0	0	0	0
Sadness	0	0	0	0	0	0	0

Table 1: Matrix of the emotion **Mild-Joy**.

7. Emotion model

We describe each emotion through a 7x7 matrix. The rows correspond to the seven recorded emotions, whereas the columns are the lip parameters. A value in the matrix represents the percentage of dependence that the corresponding lip parameter has on the corresponding emotion. Therefore, the value of the targets for each labial parameter will be an interpolation among the targets in the emotions that have a value on the column different from zero. Obviously the seven emotions have all 1 in the row corresponding to the emotion itself.

Let us consider the consonant /b/ in the triphone $/'aba/$ uttered with the emotion 'mild-joy'. The matrix of this emotion is shown in Table 1.

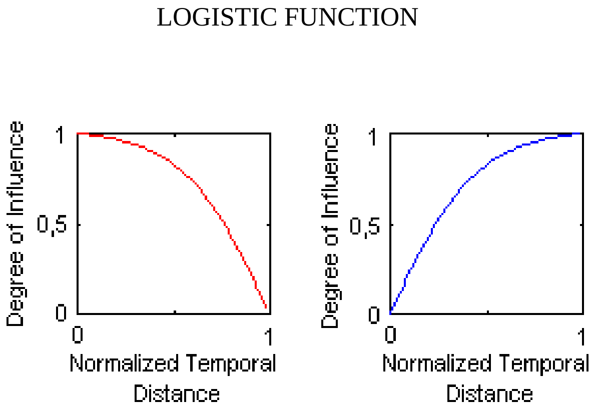
Now, let N_ab_a be the target of the lip width parameter of /b/ uttered in a neutral emotion and let J_ab_a be the target of the same lip parameter of /b/ uttered in the 'joy' emotion. The new value of this target MJ_ab_a in the mild-joy emotion will be:

$$MJ_ab_a = 0.2 * N_ab_a + 0.8 * J_ab_a$$

As consequences, the lip width will be less wide in the 'mild-joy' emotion than in the 'joy' emotion.

7.1. Expressivity qualifiers

We have also defined two qualifiers to modulate the expressivity of a lip movement. The first one, *Tension Degree*, can be *Strong*, *Normal* and *Light*. It allows one to set different intensities of muscular strain. Such a tension may appear for the expressions of emotions like fear and anger. Such a tension can also appear, for example, when a bilabial consonant (as /b/) is uttered and lips compress against each other, or when labial width increases and lips stretch them getting thinner. The second qualifier, *Articulation Degree*, can take the values *Hyper*, *Medium* and *Hypo*. During hypo articulation, it may happen that lip targets may not reach their apex.



(a) Anticipatory coarticulation.

(b) Carry-Over coarticulation.

Figure 6: Logistic Function - Strong Degree of influence.

8. Coarticulation and correlation rules

Once all the necessary visemes have been loaded from the database and modified according to the emotions and the expressiveness qualifiers, coarticulation and correlation rules are applied. In fact, to be able to represent visemes associated to vowels and consonants, we need to consider the context surrounding them [2]. Firstly, let us consider consonants. Since vowels are linked by a hierarchical relation for their degree of influence over consonants ($u > o > i > e > a$) [18, 27], we first determine which vowels will affect the consonants in $V_1C_1 \ldots C_nV_2$ sequence and which labial parameters will be modified by such an influence. Vowels act mainly over those lip parameters that characterize them. The new targets of the consonants for each lip parameter are computed through a linear interpolation between the consonantal targets ${}_{V_1}C_{iV_1}$ in the context deriving from the vowel V_1 and the consonantal targets ${}_{V_2}C_{iV_2}$ in the context of the vowel V_2:

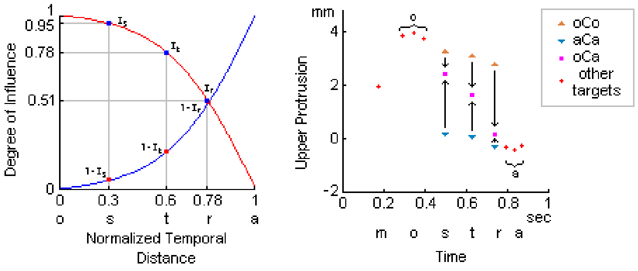
$${}_{V_1}C_{iV_2} = k * {}_{V_1}C_{iV_1} + (1-k) * {}_{V_2}C_{iV_2} \quad \mathbf{(1)}$$

The interpolation coefficient k is determined through a mathematical function, called *logistic function*, whose analytic equation is:

$$f(t) = \frac{a}{1+e^{-bt}} + c$$

This function represents the influence of a vowel over adjacent consonants, on the labial parameters that characterize it, and allows one to obtain carry-over coarticulation and anticipatory coarticulation (see Figure 7.1). The constants a and c force the function to be defined between 0 and 1 both on the abscissa and on the ordinate simplifying the computation. The constant b defines the slope of the curve that represents different degrees of influence. Time t=0 corresponds to the occurrence of V_1, and time t=1 corresponds to V_2. The consonants C_i are placed on the abscissa depending on their temporal normalized distance from the vowels.

INF E MOSTRA



(a) influence of /’o/ on /s/, /t/ and /r/.

(b) Alteration of UP targets for the Italian word /mostra/.

Figure 7: Coarticulation effects on consonants.

For example, let us consider the sequence /’ostra/ taken from the Italian word ’mostra’ (’exhibition’) and the lip parameter UP. To obtain the curve representing the upper protrusion, the consonantal targets of /s/, /t/ and /r/ in the context oCa must be calculated. The vowel /’o/ exerts a strong influence over the following three consonants and the algorithm chooses the steepest influence function. Figure 7(a) shows how the logistic function is applied to define the interpolation coefficients and Figure 7(b) shows how the targets are modified through equation (1).

Once all the necessary visemes have been calculated, correlation rules are applied modifying the value of the targets to simulate muscular tension. For example, when a bilabial consonant (as /b/) is uttered, lip compression must appear. Thus when a strong lip closing occurs the FAPs on the external boundary of the lips must be further lowered down.

Vocalic targets are modified in a very similar way, according to the consonantal context in which appear. Consonants are grouped on the base of which labial parameters they influence. For example, /b/, /m/ and /p/ will affect vowels on ULH and LLH parameters.

Finally, lip movement over time is obtained interpolating the computed visemes through Hermite Interpolation.

8.1. Speech rate

Speech rate strongly influences labial animation. At a fast speech rate, lip movement amplitude is reduced while at a slow rate it is well pronounced (lip height is wider when an open vowel occurs or lip compression is stronger when a bilabial consonant is uttered).

To simulate this effect, at a fast speech rate the value of targets points is diminished to be closer to the rest position; while at a slow speech rate, lips fully reach their targets.

9. Evaluation tests

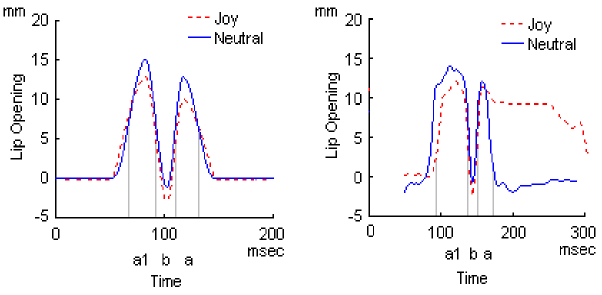
As evaluation tests, we compare the original curves with those computed by our algorithm. As first example let us consider the triphone /’aba/ uttered either in joy and in neutral expression. Figure 8 shows the curves that represent Lip Opening. We have differentiated the movement of the lower lip from the movement of the upper lip (to get a finer movement description). So the Lip Opening curves are obtained as a sum of ULH and LLH curves. The generated curves are shown in Figure 8(a) whereas the original ones in Figure 8(b). In both figures the red dotted line represents the lip movement in joy emotion, while the blue solid line describes the lip opening in neutral expression. As one can see the joy emotion causes a reduction of the lip opening (8(b)). The same behavior is also shown in generated curves (see Figure 8(a)).

The second example is quite similar to the first one but the word uttered is /mamma/ and the emotions considered are disgust and neutral. Lip Opening curves are shown in Figure 9. The red dotted line represents the lip movement in disgust emotion and the blue solid one is the lip opening in neutral expression. Figure 9(a) shows the computed curves whereas Figure 9(b) the original ones. Like joy, disgust emotion causes a diminution of lip opening. Both, the original and the computed curves display the same behavior.

In both examples, phonemes segmentation is identified by vertical lines.

For the generated curves, times are given by Festival, while for the original ones times come from the analysis of real speech. Moreover the generated curves always start at the neutral position. This is not necessarily true in the original data. Those differences can make the calculated curves slightly different from the original ones.

ABA



(a) Generated curves.

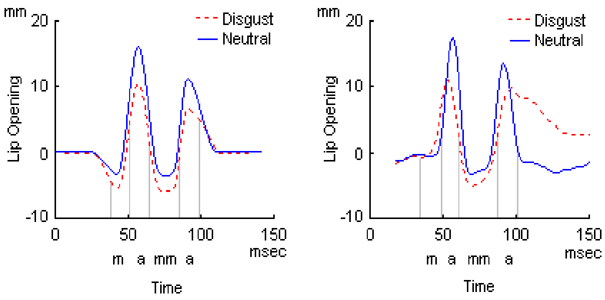
(b) Original curves.

Figure 8: Lip Opening for the triphone /’aba/ for the joy and neutral expression.

10. Conclusion and Future Development

We have presented a model that adds expressivity to the animation of an agent. Expressivity is related not only to the specification of a facial expression but also how this expression is modulated through time and depending on the context. In this paper we have presented parameters to characterize expressivity along the face modality. We have also described a computation model of emotional lip movement. Our next step is to perform some perceptual evaluation tests to further check the feasibility of our models. Movies illustrating our method may be seen at the URL: http://www.iut.univ-paris8.fr/˜pelachaud/AISB04.

MAMMA



(a) Generated curves.

(b) Original curves.

Figure 9: Lip Opening for the Italian word /mamma/ for the disgust and neutral expression.

11. References

[1] Benguerel, A. P. and Cowan, H. A., “Coarticulation of upper lip protrusion in french,” *Phonetica*, vol. 30, pp. 40–51, 1974.

[2] E. Bevacqua and C. Pelachaud. Modelling an italian talking head. In *Auditory-Visual Speech Processing AVSP’03*, Saint-Jorioz, France, 2003.

[3] M. Byun and N. Badler, “FacEMOTE: Qualitative parametric modifiers for facial animations,” *Symposium on Computer Animation*, San Antonio, TX, July 2002.

[4] D.M. Chi, M. Costa, L. Zhao, and N.I. Badler. The EMOTE model for effort and shape. In Kurt Akeley, editor, *Siggraph 2000, Computer Graphics Proceedings*, pages 173–182. ACM Press / ACM SIGGRAPH / Addison Wesley Longman, 2000.

[5] M. M. Cohen and D. W. Massaro. Modeling coarticulation in synthetic visual speech. In M. Magnenat-Thalmann and D. Thalmann, editors, *Models and Techniques in Computer Animation*, pages 139–156, Tokyo, 1993. Springer-Verlag.

[6] B. DeCarolis, C. Pelachaud, I. Poggi, and M. Steedman. APML, a mark-up language for believable behavior generation. In H. Prendinger and M. Ishizuka, editors, *Life-like Characters. Tools, Affective Functions and Applications*, pages 65–85. Springer, 2004.

[7] P. Doenges, T.K Capin, F. Lavagetto, J. Ostermann, I.S. Pandzic, and E. Petajan. MPEG-4: Audio/video

and synthetic graphics/audio for real-time, interactive media delivery, signal processing. *Image Communications Journal*, 9(4):433–463, 1997.

[8] A. Egges, S. Kshirsagar and N. Magnenat-Thalmann, "A Model for Personality and Emotion Simulation," *KES 2003*: 453-461.

[9] P. Ekman and W. Friesen. *Unmasking the Face: A guide to recognizing emotions from facial clues*. Prentice-Hall, Inc., 1975.

[10] P. Ekman and W. Friesen. Felt, false, miserable smiles. *Journal of Nonverbal Behavior*, 6(4):238–251, 1982.

[11] M. S. El-Nasr, J. Yen and T. Loerger, "FLAME - Fuzzy Logic Adaptive Model of Emotions," *International Journal of Autonomous Agents and Multi-Agent Systems*, 3(3):1-39.

[12] I. A. Essa. *Analysis, Interpretation, and Synthesis of Facial Expressions*. PhD thesis, MIT, Media Laboratory, Cambridge, MA, 1994.

[13] A. J. Fridlund and A. N. Gilbert, "Emotions and facial expression," *Science*, 230 (1985), pp. 607-608.

[14] K. Höök, A. Bullock, A. Paiva, M. Vala and R. Prada, " FantasyA - The Duel of Emotions," in *Proceedings of the 4th International Working Conference on Intelligent Virtual Agents - IVA 2003*.

[15] R. Laban and F.C. Lawrence. *Effort: Economy in body movement*. Plays, Inc, Boston, 1974.

[16] A. Löfqvist's, "Speech as audible gestures," *Speech Production and Speech Modeling*, pp. 289–322, 1990.

[17] E. Magno-Caldognetto, C. Zmarich, and P. Cosi. Coproduction of speech and emotion. In *ISCA Tutorial and Research Workshop on Audio Visual Speech Processing, AVSP'03*, St Jorioz, France, September 4th-7th 2003.

[18] E. Magno-Caldognetto, C. Zmarich and P. Cosi, "Statistical definition of visual information for Italian vowels and consonants," in *International Conference on Auditory-Visual Speech Processing AVSP'98*, D. Burnham, J. Robert-Ribes, and E. Vatikiotis-Bateson, Eds., Terrigal, Australia, 1998, pp. 135–140.

[19] S. Marsella, W.L. Johnson, and K. LaBore. Interactive pedagogical drama. In *Proceedings of the 4th International Conference on Autonomous Agents*, Barcelona, Spain, June 2000.

[20] S. Marsella and J. Gratch. Modeling coping behavior in virtual humans: Don't worry, be happy. In *proceedings of the /2nd International Conference on Autonomous Agents and Multiagent Systems*, Melbourne, Australia, 2003.

[21] D. Massaro. *Perceiving Talking Faces : From Speech Perception to a Behavioral Principle*. Bradford Books Series in Cognitive Psychology. MIT Press, 1997.

[22] C. Pelachaud, N.I. Badler, and M. Steedman. Generating facial expressions for speech. *Cognitive Science*, 20(1):1–46, January-March 1996.

[23] C. Pelachaud, V. Carofiglio, B. de Carolis, F. de Rosis and I. Poggi, "Embodied Contextual Agent in Information Delivering Agent," in *Proceedings of AAMAS*, 2002, vol. 2.

[24] I. Poggi. Mind markers. In C.Mueller and R.Posner, editors, *The Semantics and Pragmatics of Everyday Gestures*. Berlin Verlag Arno Spitz, Berlin, 2001.

[25] A. Takeuchi and K. Nagao, "Communicative facial displays as a new conversational modality," *ACM/IFIP INTERCHI '93*, Amsterdam, 1993.

[26] P. Taylor, A. Black, and R. Caley. The architecture of the Festival Speech Synthesis System. In *Proceedings of the Third ESCA Workshop on Speech Synthesis*, pages 147–151, 1998.

[27] E.F. Walther, *Lipreading*, Nelson-Hall, Chicago, 1982.

[28] Y. Yacoob and L. Davis. *Computer Vision and Pattern Recognition Conference*, chapter Computing spatio-temporal representations of human faces, pages 70–75.