Empirical investigation of the temporal relations between speech and facial expressions of emotion

Stéphanie BUISINE [1*], Yun WANG [1&2] & Ouriel GRYNSZPAN [3]

1. Arts et Métiers ParisTech, LCPI, 151 bd de l'Hôpital, 75013 Paris, France

2. LIMSI-CNRS, BP 133, 91403 Orsay Cedex, France

3. CNRS USR 3246, Université Pierre et Marie Curie, Hôpital de la Salpêtrière, 47 bd de l'Hôpital, 75013 Paris, France

* Corresponding author: Phone +33.144.246.377 - Fax: +33.144.246.359 - stephanie.buisine@ensam.eu

Abstract: Behavior models implemented within Embodied Conversational Agents (ECAs) require nonverbal communication to be tightly coordinated with speech. In this paper we present an empirical study seeking to explore the influence of the temporal coordination between speech and facial expressions of emotions on the perception of these emotions by users (measuring their performance in this task, the perceived realism of behavior, and user preferences). We generated five different conditions of temporal coordination between facial expression and speech: facial expression displayed before a speech utterance, at the beginning of the utterance, throughout, at the end of, or following the utterance. 23 subjects participated in the experiment and saw these 5 conditions applied to the display of 6 emotions (fear, joy, anger, disgust, surprise and sadness). Subjects recognized emotions most efficiently when facial expressions were displayed at the end of the spoken sentence. However, the combination users viewed as most realistic, preferred over others, was the display of the facial expression throughout speech utterance. We review existing literature to position our work and discuss the relationship between realism and communication performance. We also provide animation guidelines and draw some avenues for future work.

Keywords: Temporal coordination, facial expression, emotion, perception.

1. Introduction

Embodied Conversational Agents (ECAs) are interactive virtual characters which take on a communicative role in various application fields (e.g. e-learning, games, e-commerce, therapeutic tools) using several modality channels such as speech, facial expressions, gestures, postures, etc. The ideal ECA [16] should be intelligent, capable of social behavior, and should take advantage of its visual representation to strengthen its believability (particularly by means of sophisticated and relevant nonverbal behavior, and by the expression of

emotions). The notion of believability is a central one in ECA research: it relies mainly on the visual properties of the agent and on the generation of verbal and nonverbal behavior during interaction with the user [17]. It is nonetheless a complex notion tying back to the concepts of *naturalism* or realism of agent behavior and effectiveness of communication. Yet past research has suggested that realism does not always correlate to communication effectiveness: studies by Calder and colleagues showed that caricaturing facial expressions, although this decreased ratings of human-likeness or plausibility, increased recognition of agents' emotions by subjects (shorter reaction times), increased their neural response and ratings of emotional intensity [3]. This paper focuses on the temporal arrangement of speech and facial expressions of emotions within the believability framework, addressing both the issues of realism and of effectiveness. We first review the literature related to the coordination of speech and facial expressions (Section 2) and justify an empirical exploratory procedure to extend our knowledge. Following this, we describe our methods and results (Section 3) and discuss the implications for theory and design of ECA (Section 4).

2. Coordination of speech and facial expressions

Research related to the generation of ECAs' nonverbal behaviors stresses the importance of defining their temporal coordination speech-based communication. One challenge for ECA platforms is to control very precisely the synchronization of communication channels [13]. In terms of software architecture, this implies simultaneous generation of these various communication channels from a unique representation (e.g. facial expressions should not be derived from the speech content but must be generated simultaneously). BEAT (Behavior Expression Animation Toolkit, [6]) is an example of a framework allowing the automatic generation of animations synchronizing speech synthesis, voice intonation, eyebrow movements, gaze direction, and hand gestures.

From a functional standpoint [29], facial expressions can take on *semantic* (e.g. to emphasize or substitute for a word), *syntactic* (e.g. nodding, raising eyebrows to emphasize parts of the speech flow), *dialogic* (e.g. gazes to regulate speech turns) or *pragmatic* (e.g. expressing the speaker's personality, emotions or attitudes) functions in a conversation. Rules for coordination of facial expression with speech depend on these functions [19]: syntactic facial cues must coordinate with

the elementary phonemes of speech [4, 18], whereas semantic and dialogic cues are synchronized with complete words or pauses [5, 27]. Finally, pragmatic cues are synchronized with complete sentences [21] or with speech turns [7, 28], since emotions are not expected to fluctuate at the level of individual words. Therefore, there seems to be a consensus in the literature on a *synchrony paradigm*, in which facial expressions of emotion are synchronized with and displayed throughout speech. Related research has investigated the interaction of facial expressions of emotion with other facial cues (lip-synching, or facial expressions with an alternate function) and set specific priority rules and additive rules, as well as methodologies for conflict resolution [28]. Platforms were designed to support dynamic representations of emotions [26], in particular to replicate realistic emotional control (e.g. related to an agent's behavior or mood [9, 31]). Some aspects of emotion dynamics were submitted to perceptive tests, in particular the onset, offset and apex durations of facial expressions of emotion [20]. It was shown for example that slow onset smiles lead to more positive perceptions (e.g. smiles are perceived as more authentic, and the person as more attractive and trustworthy).

Although the synchrony paradigm seems largely undisputed in the aforementioned literature, there are at least three reasons to question it in our view. Firstly, from an engineering viewpoint, it is imprecise and does not provide sufficient guidelines to help position emotional expressions (i.e. emotional tags) in relation to speech utterances. In this respect our goal is to refine animation rules in order to reduce the role of chance or of the animator's own talent. Secondly, according to the sequential checking process of appraisal [30], emotions may not all have the same dynamics, and different emotions might support different intensity patterns in the course of speech. For example, given that 1) surprise is assumed to rely on novelty appraisal and anger on goal-conduciveness appraisal, and 2) that novelty appraisal is supposed to occur earlier than goal-conduciveness appraisal in the checking process, one can hypothesize that surprise and anger dynamics are different. Accordingly, surprise would yield an earlier intensity peak than anger. Finally, we wished to investigate alternative coordination patterns in search of more effective communication. As previously mentioned, stylized or caricatured animations (with stereotypical behaviors and mental states, conveyed transparently) can be more efficient than ecological or naturalistic communication

styles [3, 12]. Therefore, the present study aims to challenge the synchrony paradigm.

3. Experiment

3.1. Goal

Our general motivation is to explore the effects of various temporal coordination rules between speech and facial emotional expressions on communication effectiveness and realism, and to examine their interaction with several emotions. We set up 5 different temporal combinations of speech and facial expressions of emotions, implemented them within an ECA platform and applied them to the display of 6 fundamental emotions. By means of a perceptive test, we assessed their impact on three criteria: the effectiveness of communication (ability to convey the intended emotion, i.e. recognition score, perceived intensity, answer time), its perceived realism, and user preferences (subjective criteria).

3.2. Methods

3.2.1. Participants and material

23 subjects participated in the experiment (17 men and 6 women, 25.4 years old on average, SD=5.7) including engineering students, artists, and administrative staff members, all native speakers of French.

We chose to investigate Ekman's six fundamental emotions [11] (Joy, Disgust, Sadness, Surprise, Fear, and Anger) because they are well-documented and known to be universally perceived [10]. The related facial expressions were generated in Poser (http://my.smithmicro.com/win/poser/) following Ekman's review guidelines [11]. We implemented one expression for each emotion and chose a congruent short sentence to be associated to each of them: Joy was associated to "I watched my favorite program", Sadness to "I have to work all weekend", Disgust to "We're being served spinach", Surprise to "My train is 20 minutes late", Fear to "I've lost my father's phone", and Anger to "Charles has hidden my book". The sentences were chosen so that their meaning would not be straightforward, thus leaving minimal uncertainty that the facial expression allowed to resolve. Conversely, the sentence had to provide minimal cues for decoding the facial expression, given that our goal is to investigate how

modalities cooperate with emotion decoding. All sentences were synthesized in French with voice intonation set to neutral, using Acapela's Virtual Speaker text-to-speech (http://www.acapela-group.com/) and GoldWave audio editor (http://www.goldwave.com/).

The integration of speech and facial expression was performed with Poser: lip-synching was generated automatically, and the facial expressions were inserted manually following 5 patterns inspired from Allen's typology of temporal relations [1]. Table 1 describes these patterns by representing only the most expressive stage of facial expressions (the apex, or sustain), which represented approximately 30% of the utterance duration (m=42.3 frames and 1.41 sec), except for the "during speech" condition, in which the apex covered approximately 60% (m=81.7 frames and 2.72 sec) of utterance duration. Attack and decay phases (the onset and offset), not represented in Table 1, lasted 611 ms each on average (m=18.34 frames), which corresponds to "slow" onsets and offsets [20]. The "during speech" condition corresponds to the aforementioned synchrony paradigm (in which the facial expression covers the whole utterance).

Name	**Illustration**
Before speech	Facial expr. / Spoken utterance
Beginning of speech	Spoken utterance / Facial expr.
During speech	Spoken utterance / Facial expression
End of speech	Spoken utterance / Facial expr.
After speech	Spoken utterance / Facial expr.

Table 1. The five temporal patterns tested in the experiment.

The resulting 30 animations (5 temporal patterns for each of the 6 target emotions) comprised 110 to 150 frames, depending on the length of the spoken sentence and on the combination pattern chosen, with an average duration of 4.55 seconds in total.

3.2.2. Procedure

We used a full within-subject factorial design with 2 stages. In the first stage, each subject had to successively examine the 30 animations in a random order, label each one with the emotion perceived (in the subject's own words) and rate its

perceived intensity on a 7-point Likert scale (see Figure 1 left panel). In completing this exercise the subject could replay each animation as many times as necessary.

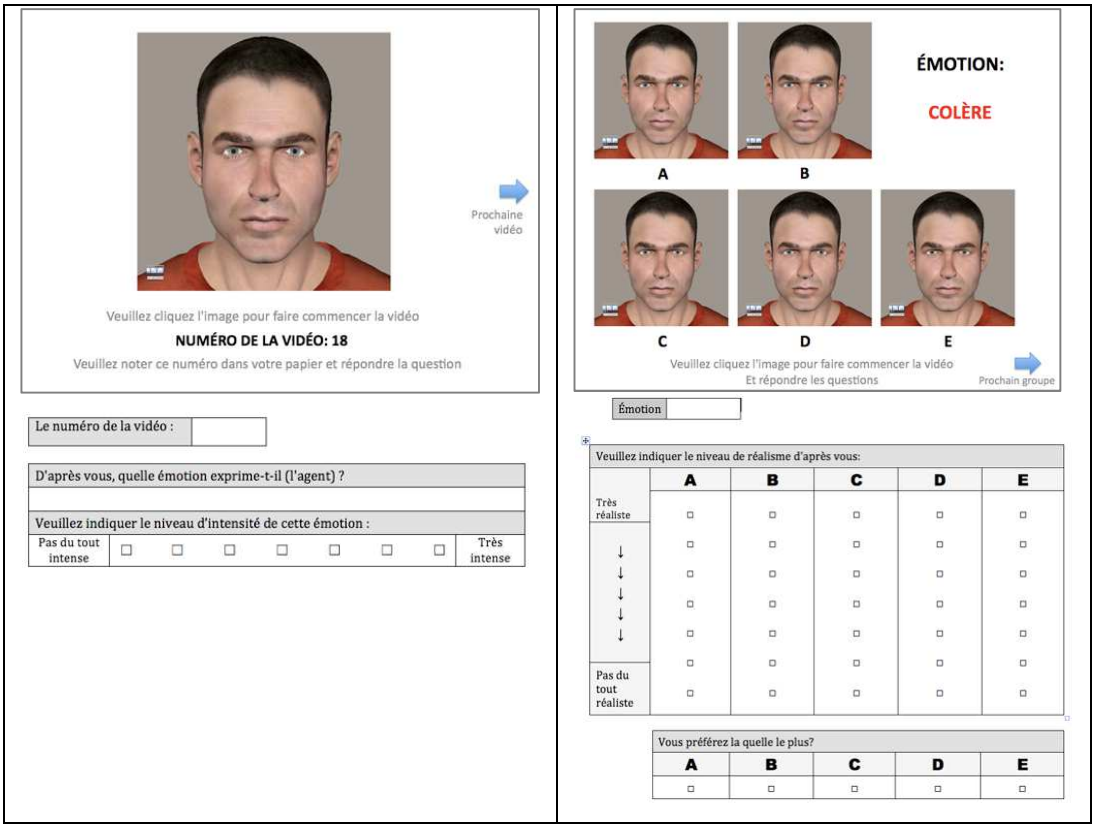

Figure 1: Left panel: test interface for the first stage: animations are randomly presented one by one, and subjects label the perceived emotion and rate its perceived intensity. Right panel: interface for the second stage: blocks of 5 animations are associated to the same target emotion; subjects rate perceived realism and express their preference).

In the second stage, the 5 animations generated for a given emotion (corresponding to the 5 temporal patterns) were gathered in a single display (the arrangement of the 5 animations within the display was randomized) and the intended emotion was brought to the subject's notice (see right panel in Figure 1). Subjects rated the realism of each animation using a 7-point Likert scale, and chose their preferred animation out of the 5 displayed.

3.2.3. Data collection and analysis

The recognition of the target emotion for each animation was scored as true (1) or false (0) by 2 independent judges by examining the emotion label or labels which subjects attributed to the various animations. The judges first obtained 92.03% agreement. To form the final dataset, discrepancies were solved by consensus between the judges. For example, when several labels were used for a single

animation, one judge tended to consider that recognition was achieved as far as one label was relevant, whereas the second judge considered such answers as wrong as far as one label was irrelevant. We finally adopted the latter, more conservative, rule in order to prevent a ceiling effect and maximize the likelihood of observing differences between our various experimental conditions.

The duration of each trial (time used to view the animation, label the emotion, rate perceived intensity and validate the trial) was also recorded as an indirect index of ease of recognition. The other variables collected were perceived intensity of emotions, perceived realism (1 to 7 scores) and the preference ratings (1 or 0 for each animation). Data were analyzed by means of ANOVAs with Emotions and Temporal Patterns as within-subject variables. Fisher's LSD was used for post-hoc pairwise comparisons. Moreover, linear correlation analyses were performed on the whole set of dependent variables (recognition, trial time, perceived intensity, realism and preference).

3.3. Results

3.3.1. Decoding performance: recognition, response time, and perceived intensity

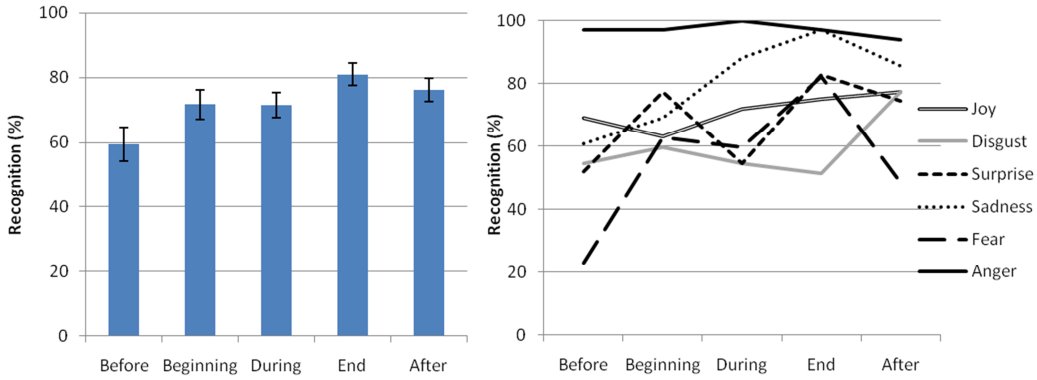

Figure 2: Effect of Temporal Patterns on recognition rate (left panel) and interaction between Temporal Patterns and Emotions (right panel).

The overall average recognition score amounted to 72%. The main effect of Emotions on this performance proved to be significant ($F(5/105)=4.84$; $p=0.001$; $\eta^2=0.187$), with Anger being significantly better recognized than were all other emotions ($p<0.017$). Temporal Patterns also significantly influenced recognition performance ($F(4/84)=5.92$; $p<0.001$; $\eta^2=0.220$; see left panel in Figure 2). The "before speech" condition yielded significantly lower recognition rates than all

other conditions ($p<0.025$). The most effective conditions were “end of speech” and “after speech”. In particular, it is worth mentioning that subjects were significantly more effective in the “end of speech” than in the “during speech” condition ($p=0.05$).

An interaction was also observed between Emotions and Temporal Patterns ($F(20/420)=2.41$; $p<0.001$; $\eta^2=0.103$; see right panel in Figure 2), showing that the influence of the Temporal Patterns varied depending on the emotion expressed: for example recognition rate for Anger was always high and did not depend on the Temporal Pattern displayed, whereas recognition rate for Fear varied between 23% (“before speech”) and 82% (“end of speech” condition). Subject gender did not affect the recognition rates ($F(1/21)=1.19$; NS).

Response time was analyzed on a subsample of 22 subjects, since the recording procedure failed for one of the subjects. Average response time for a trial in the first experimental stage was 25 sec. We observed only a main effect of Emotion ($F(5/100)=3.60$; $p=0.005$; $\eta^2=0.152$) for this variable: for example response time for Disgust and Sadness was lower than for Joy, Surprise and Fear ($p<0.07$).

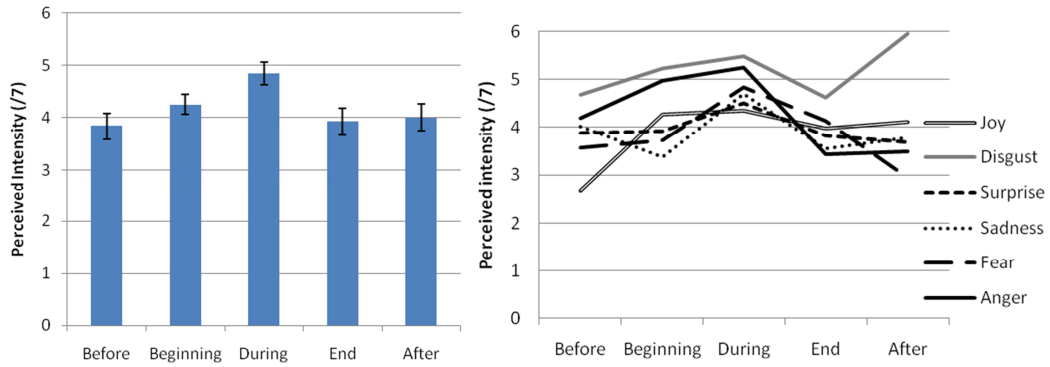

Figure 3: Effect of Temporal Patterns on the perceived intensity of emotions (left panel) and interaction between Temporal Patterns and Emotions (right panel).

As for perception of intensity, a main effect of Emotions was observed ($F(5/105)=9.19$; $p<0.001$; $\eta^2=0.299$): the expression of Disgust was perceived as significantly more intense than that of the other emotions ($p<0.001$), which did not significantly differ from one another. Temporal Patterns also influenced the perceived intensity of emotions ($F(4/84)=5.93$; $p<0.001$; $\eta^2=0.216$; see Figure 3 left panel): the “during speech” condition resulted in significantly higher perceived intensities than did other temporal pattern conditions ($p<0.011$), which did not differ significantly from one another. There was an interaction between Emotions and Temporal Patterns ($F(20/420)=2.89$; $p=0.001$; $\eta^2=0.106$; see right

panel in Figure 3), suggesting that the influence of Temporal Patterns was not constant over all emotions: for example the “after speech” condition resulted in more intense perception of Disgust but tended to decrease the perceived intensity of other emotions.

3.3.2. Realism and preferences

Scores for perceived realism vary significantly depending on the Emotions ($F(5/105)=2.79$; $p=0.021$; $\eta^2=0.117$): for example, the expression of Disgust obtained higher scores of realism than did expressions of Joy and Fear ($p<0.07$). Temporal patterns also influenced perceived realism ($F(4/84)=24.19$; $p<0.001$; $\eta^2=0.535$; see Figure 4 left panel): the condition perceived as most realistic was “during speech” ($p<0.001$) and the one perceived as least realistic was “after speech” ($p<0.008$). An interaction between Emotions and Temporal Patterns ($F(20/420)=3.45$; $p<0.001$; $\eta^2=0.141$; see Figure 4 right panel) shows minor variations in this result.

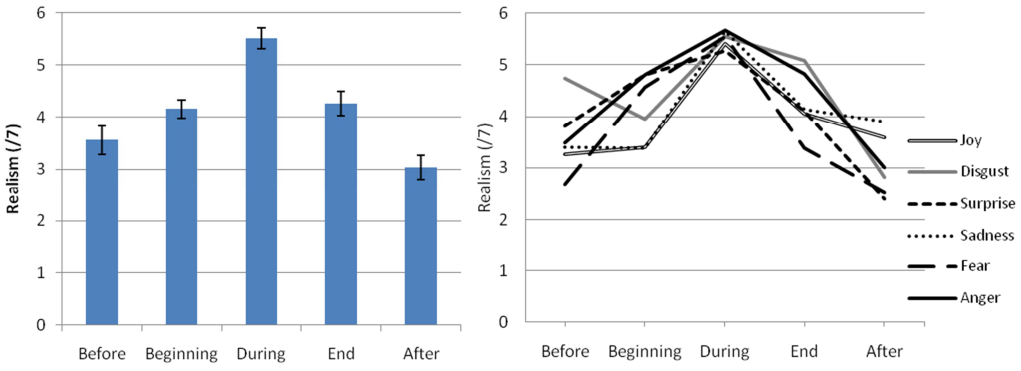

Figure 4: Effect of Temporal Patterns on the perceived realism of animations (left panel) and interaction between Temporal Patterns and Emotions (right panel).

Temporal Patterns also significantly influenced subjects’ preferences ($F(4/84)=25.87$; $p<0.001$; $\eta^2=0.552$): the “during speech” condition was preferred to all others ($p<0.001$). An interaction between Emotions and Temporal Patterns ($F(20/420)=1.87$; $p=0.013$; $\eta^2=0.082$) showed that these preferences are more or less strict depending on the displayed emotion: for example the “during speech” condition was largely preferred for the expression of Sadness and less strongly so for the expression of Surprise, which was also well rated in terms of preference, in both the “before speech” and “beginning of speech” conditions.

3.3.3. Correlations between variables

To complete our data analysis we performed pairwise linear correlation tests on our five dependent variables (see Table 2). Results show that the recognition score is negatively correlated to response time and does not correlate with any other variable. Intensity and realism correlated negatively to answer time; perceived intensity correlated positively to both realism and subjective preference ratings. Finally, preference and realism were exhibited a strongly positive correlation.

Answer time	-0.2			
Intensity	*(-0.01)*	-0.31		
Realism	*(0.09)*	-0.32	**0.53**	
Preference	*(-0.02)*	*(-0.07)*	**0.5**	**0.8**
	Recognition	Answer time	Intensity	Realism

Table 2: Linear correlation coefficients between our dependent variables. Weak correlations that can be considered as null ($|r|<0.2$) are in parentheses and italics, medium correlations ($0.2<|r|<0.4$) are in normal font, strong ($0.4<|r|<0.6$) and very strong ($|r|>0.6$) correlations are in bold font.

4. Discussion and conclusion

Some of our results, for example the average recognition score of 72%, provide indirect validation of our designs of facial expressions. Anger was particularly well decoded and our expression of disgust was perceived as globally more intense (and more realistic) but the other stimuli constituted a homogeneous corpus. The differences between our recognition rates and those obtained by Ekman [10] can be explained by several facts. Firstly, we used animations instead of static images. Secondly, synthetic characters caricature human features. This could have allowed for superior recognition rates (e.g. anger) when compared with real pictures as in Ekman's work. Conversely, imperfections in some of our models of facial expressions and/or in graphical rendering, as well as the specific format of our protocol (free response instead of forced-choice format) could explain why other recognition rates (e.g. for joy) were inferior to Ekman's.
In any case, the most effective combination of speech and emotional facial expression consisted in positioning the facial expression at the end of speech utterances. This animation pattern significantly outperformed (with a 10% improvement of recognition) the more realistic synchrony paradigm, which is an unexpected and interesting result. Closer examination of interactions between

temporal animation patterns and emotions, suggests that the synchrony (or "during speech") condition gave rise to poor recognition scores (i.e. lower than the average recognition rate) for expressions of fear, surprise, and disgust (see Table 3 for a summary of our design recommendations to enhance recognition of emotions by users). Furthermore, the expression of disgust is a noteworthy exception to the good performance of the "end of speech" condition: disgust was much better recognized (+26%) when displayed after the speech. The expression of disgust involves special movements of the lips which are important to distinguish it from anger (which shares some features of the nose, eyebrows and eyes expressions with disgust) and from the "clueless" state (which share features of the eyes and eyebrows with disgust [24]). This may explain why disgust is better recognized when the lip movements are dedicated to the expression of emotions, i.e. in the absence of speech articulation.

Emotion:	**To be favored:**	**To be avoided:**
Anger	During speech	
Disgust	After the speech	At the end of or during speech
Fear	At the end of speech	During or before speech
Joy	After the speech and/or accented at the end	At the beginning of speech
Sadness	Accented at the end of speech	Before or at the beginning
Surprise	At the beginning of speech (preferably) or at the end of speech	During speech

Table 3: Design recommendations to position facial expressions of emotion with respect to speech utterance in ECAs in order to increase recognition of emotions by their users.

The temporal pattern also influenced the perceived intensity of emotions, but in a different way: the "during speech" condition resulted in higher perceived intensity, which can be attributed to the total duration of the facial expression – in the "during speech" condition the facial expression was twice as long as in all other conditions – and is consistent with existing literature suggesting that duration and intensity are correlated in the generation [23] and perception [25] of facial expressions. However, variations in the duration of facial expressions constitute a limitation of our study and would require a new experimental iteration in order to be better understood: in particular, we wonder whether they might introduce biases in the evaluation of realism since the correlation matrix showed that realism was strongly correlated to intensity of the stimuli.

Conversely, the fact that other temporal patterns did not significantly interfere with the perception of emotion intensity is a positive finding and suggests that

flexibility in encoding emotion intensity is preserved when manipulating the temporal coordination of speech and facial expressions of emotion.

The synchrony or "during speech" condition showed significantly higher realism and preference scores than all other conditions. Hence one should adapt the animation strategy according to the primary goal of the application since one cannot optimize both realism and decoding performance at the same time. A tradeoff can be met with a "during+after" coordination pattern: we have chosen this solution in designing a tool for socio-cognitive training for people with autism, in which both recognition effectiveness and realism were important [15].

Correlations observed between our dependent variables also indirectly validate the global consistency of our dataset: the stimuli which were processed faster were better recognized, rated as more intense, and more realistic. However, the strong correlation between realism and user preferences opens the discussion related, for example, to the "Uncanny Valley" theory which proved particularly well suited to model the realism of agent behavior [14]. This theory predicts that agents demonstrating high realism might be less well evaluated than agents demonstrating only moderate realism, which is inconsistent with our results. To explain this discrepancy we hypothesize that our agents have not reached the valley boundaries in this experiment, since we used a neutral speech intonation with facial expressions of emotions in an emotional context. The average realism score obtained by our animations ($m=4.1/7$, $\sigma=0.2$) might therefore still position our agents as being "moderately realistic" explaining the positive evaluations they received. Such hypothesis opens up avenues for new experimental investigations: one of the first steps in our future work will consist in introducing emotional speech prosody [8], diversifying the sentences associated to each emotion and replicating the present protocol. Other future directions will be to include longer sentences and the expression of mixed emotions [2, 22], which may surely raise new challenges for animation and perception of emotions in ECAs.

Acknowledgements

This study was supported by a grant from La Fondation de France and La Fondation Adrienne et Pierre Sommer, as part of a wider project on Virtual Environments for Socio-Cognitive Training in Autism (project EVESCA, Engt n°2007 005874, coordinated by Ouriel Grynszpan). The authors thank Jean-Claude Martin and Julien Nelson for their contributions.

References

[1] Allen, J. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26, pp. 832-843.

[2] Arya, A., DiPaola, S., & Parush, A. (2009). Perceptually valid facial expressions for character-based applications. *International Journal of Computer Games Technology*, pp. 1-13.

[3] Calder, A.J., Rowland, D., Young, A.W., Nimmo-Smith, I., Keane, J., & Perrett, D.I. (2000). Caricaturing facial expressions. *Cognition*, 76, pp. 105-146.

[4] Cassell, J. (2000). Nudge nudge wink wink: Elements of face-to-face conversation for embodied conversational agents. In J. Cassell, J. Sullivan, S. Prevost & E. Churchill (Eds.), *Embodied Conversational Agents*, pp. 1-27. Cambridge: MIT Press.

[5] Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S., & Stone, M. (1994). Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. SIGGRAPH'94, pp. 413-420, ACM Press.

[6] Cassell, J., Vilhjálmsson, H., & Bickmore, T. (2001). BEAT: the Behavior Expression Animation Toolkit. SIGGRAPH '01, pp. 477-486,

[7] De Rosis, F., Pelachaud, C., Poggi, I., Carofiglio, V., & De Carolis, B. (2003). From Greta's mind to her face: Modelling the dynamics of affective states in a conversational embodied agent. *International Journal of Human-Computer Studies*, 59, pp. 81-118.

[8] Devillers, L., Vidrascu, L., & Lamel, L. (2005). Emotion detection in real-life spoken dialogs recorded in call center. *Journal of Neural Networks*, 18, pp. 407-422.

[9] Egges, A., Kshirsaga, S., & Magnenat-Thalmann, N. (2004). Generic personality and emotion simulation for conversational agents. *Computer Animation and Virtual Worlds*, 15, pp. 1-13.

[10] Ekman, P. (1994). Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique. *Psychological Bulletin*, 115, pp. 268-287.

[11] Ekman, P., & Friesen, W.V. (1975). *Unmasking the face. A guide to recognizing emotions from facial clues.* Englewood Cliffs, New Jersey: Prentice-Hall.

[12] Gratch, J., Marsella, S., Egges, A., Eliëns, A., Isbister, K., Paiva, A., Rist, T., & ten Hagen, P. (2004). *Design criteria, techniques and case studies for creating and evaluating interactive experiences for virtual humans.* Working group on ECA's design parameters and aspects, Dagstuhl seminar on Evaluating Embodied Conversational Agents.

[13] Gratch, J., Rickel, J., André, E., Badler, N., Cassell, J., & Petajan, E. (2002). Creating interactive virtual humans: Some assembly required. *IEEE Intelligent Systems*, 17, pp. 54-63.

[14] Groom, V., Nass, C., Chen, T., Nielsen, A., Scarborough, J.K., & Robles, E. (2009). Evaluating the effects of behavioral realism in embodied agents. *International Journal of Human-Computer Studies*, 67, pp. 842-849.

[15] Grynszpan, O., Nadel, J., Constant, J., Le Barillier, F., Carbonell, N., Simonin, J., Martin, J.C., & Courgeon, M. (2009). A new virtual environment paradigm for high functioning autism

intended to help attentional disengagement in a social contexte. Virtual Rehabilitation International Conference, pp. 51-58,

[16] Isbister, K., & Doyle, P. (2004). The blind men and the elephant revisited. In Z. Ruttkay & C. Pelachaud (Eds.), *From brows to trust: Evaluating Embodied Conversational Agents.*, pp. 3-26: Kluwer Academic Publishers.

[17] Johnson, W.L., Rickel, J., & Lester, J. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education*, 11, pp. 47-78.

[18] Krahmer, E., & Swerts, M. (2004). More about brows. In Z. Ruttkay & C. Pelachaud (Eds.), *From brows to trust: Evaluating Embodied Conversational Agents*, pp. 191-216: Kluwer.

[19] Krahmer, E., & Swerts, M. (2009). Audiovisual prosody - Introduction to the special issue. *Language and Speech*, 52, pp. 129-133.

[20] Krumhuber, E., Manstead, A.S.R., & Kappas, A. (2007). Temporal aspects of facial displays in person and expression perception: The effect of smile dynamics, head-tilt, and gender. *Journal of Nonverbal Behavior*, 31, pp. 39-56.

[21] Lester, J., Towns, S., Callaway, C., Voerman, J., & FitzGerald, P. (2000). Deictic and emotive communication in animated pedagogical agents. In J. Cassell, S. Prevost, J. Sullivan & E. Churchill (Eds.), *Embodied Conversational Agents*, pp. 123-154. Cambridge: MIT Press.

[22] Martin, J.C., Niewiadomski, R., Devillers, L., Buisine, S., & Pelachaud, C. (2006). Multimodal complex emotions: Gesture expressivity and blended facial expressions. *International Journal of Humanoid Robotics*, 3, pp. 269-292.

[23] Messinger, D.S., Fogel, A., & Dickson, K.L. (1999). What's in a smile? *Developmental Psychology*, 35, pp. 701-708.

[24] Nusseck, M., Cunningham, D.W., Wallraven, C., & Bülthoff, H.H. (2008). The contribution of different facial regions to the recognition of conversational expressions. *Journal of Vision*, 8, pp. 1-23.

[25] Pelachaud, C. (2005). Multimodal expressive embodied conversational agents. International Multimedia Conference, pp. 683-689,

[26] Pelachaud, C. (2009). Modelling multimodal expression of emotion in a virtual agent. *Philosophical Transactions of the Royal Society B*, 364, pp. 3539-35648.

[27] Pelachaud, C., Badler, N., & Steedman, M. (1996). Generating facial expressions for speech. *Cognitive Science*, 20, pp. 1-46.

[28] Pelachaud, C., Carofiglio, V., De Carolis, B., de Rosis, F., & Poggi, I. (2002). Embodied contextual agent in information delivering application. AAMAS'2002, pp. 758-765,

[29] Scherer, K.R. (1980). The functions of nonverbal signs in conversation. In H. Giles & R. St Clair (Eds.), *The Social and Physhological Contexts of Language*, pp. 225-243: LEA.

[30] Scherer, K.R. (2001). Appraisal considered as a process of multi-level sequential checking. In K.R. Scherer, A. Schorr & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, Methods, Research*, pp. 92-120. New York and Oxford: Oxford University Press.

[31] Tanguy, E., Willis, P., & Bryson, J. (2007). Emotions as durative dynamic state for action selection. IJCAI'07 International Joint Conference on Artificial Intelligence, pp. 1537-1542,