

# AMUSE: A Tool for Evaluating Affective Interfaces

**Noël Chateau**

France Telecom R&D Division  
2, Avenue Pierre-Marzin  
22307 Lannion, France  
noel.chateau@francetelecom.com  
+33 2 96 05 32 42

**Marc Mersiol**

France Telecom R&D Division  
2, Avenue Pierre-Marzin  
22307 Lannion - France  
marc.mersiol@francetelecom.com  
+33 2 96 05 19 04

## ABSTRACT

The design of affective interfaces introduces new challenges for their assessment. New methods and tools have to be developed, taking into account the emotional dimension of Computer-Human Interactions. This paper presents the tool AMUSE that allows to conduct such evaluations by collecting and aggregating different sources of data including user's eye gaze and physiological data. Some results of an experiment on simulated Embodied Conversational Agents (ECAs) are briefly presented as an illustration of the new possibilities of investigation brought by AMUSE.

## Author Keywords

Human factors, evaluation, emotion, affective interfaces, Embodied Conversational Agents, integrated tool.

## ACM Classification Keywords

HCI evaluation.

## INTRODUCTION

On growing competitive markets where more and more products and services are proposed for roughly equivalent prices, the conditions of their successful adoption by customers do not lie any more only on their utility and usability, but also on the pleasure they give [6]. The "affective design" of "pleasurable products and services" (see the special issue on this topic in [4]) has become a focus of many companies aiming at incorporating the emotional dimension in their design process.

This evolution of design opens new methodological questions for the evaluation of users' perceptions and reactions to these products and services. Indeed, up to now, user tests conducted during the design process were mainly dedicated to assessing the perceived quality and the ergonomics of the product or service under study. Today, new methods and tools are emerging as first answers to the need for assessing the emotional dimension of interactions between users and products or services. As human emotional processes are linked to perception, cognition and action, those methods and tools inherit from a wide range of existing ones, mainly coming from (electro- and neuro-) physiology, psychophysics, psychology, cognitive ergonomics, linguistics, communication, marketing and sociology. Some proposed methods and tools record

instantaneous signals from the user (skin conductance, heart rate, facial mimics, gesture, voice) during the test and require signal processing for further analysis [10]. In that case, the focus is put on user's instantaneous emotional reactions, but nothing is known on his/her own perception of his/her emotional state. Other methods and tools directly ask the user to conduct an *a posteriori* self-assessment of the emotion he/she experienced during the test by using questionnaires (e.g., [8]) or pictorial representations of him/her-self (e.g., Self-Assessment Manikin from [7]; PrEmo from [3]). Here, the focus is put on the user's own opinion which might be richer than the signs he/she emitted during the test, but one can only access this opinion through the filter of memory and language or graphics. For a thorough understanding of users' emotional reactions during a test, it might then certainly be beneficial to combine both instantaneous and *a posteriori* methods.

When instantaneous signals are recorded during a test, the first major issue lies in the synchronization and fusion of all possible sources of information. Indeed, electro-physiologic, video, audio or any other sensors may work at different sample rate and provide data of different formats. These data need to be combined in one global matrix for further processing and statistical analysis. The second major issue rests with the amount of collected data to be analyzed. For example, if one uses six sensors working at a mean sample rate of 10 Hz (one scalar data captured every 100 ms), a 30-min test will result in six vectors of 18000 scalar values each to jointly analyze. For an efficient analysis, it is therefore crucial to furnish tools to the experimenter that allow a simple and comprehensive visual representation of all of these data.

This paper presents a tool call AMUSE that was developed at France Telecom's R&D lab for conducting evaluations of affective interfaces proposed on Personal Computers (PCs). This tool offers the possibility to record and synchronize signals from eight electro-physiologic sensors, an eye-tracker, a mouse and keyboard tracker, windows displayed on the computer, a video of the user and to mark any event of interest for the experimenter. A visual display of the recorded data in a "replay mode" allows the experimenter to subsequently conduct a refined analysis of the test progress. A brief example of use of AMUSE in a test on simulated ECAs will be also presented.

## DESCRIPTION OF THE TOOL AMUSE

### Equipment and collected data

AMUSE is made up of three PCs: a server, a PC dedicated to the eye-tracker and a client used by the user participating to the test. The server is dedicated to the creation of the test and to data collection. The eye-tracker PC collects the data of the eye-tracking. The client records data of the interaction with the user. All PCs are connected within a local network through a TCP/IP protocol. AMUSE is designed in a modular way. Four modules allow to collecting, synchronizing and merging four categories of data.

### Data concerning the user's eye gaze on a computer screen

AMUSE enables the experimenter to follow without interruption a user's eye gaze in a non intrusive manner by using an eye-tracker. Several data can be analysed such as: gaze position, fixation number, fixation duration, repeat fixations, search patterns, etc. From these data potential cognitive measures can be inferred, for example, difficulty, attention, stress, relaxation, successful problem solving, higher level of reading skill [13].

### Data concerning the interaction with a PC interface

AMUSE automatically collects events characterizing user's interaction with a system for example when surfing on a Web site or when using any Windows® application. These events are mouse clicks, struck keys, new URLs or windows opened, etc.

### Data concerning user's activity

AMUSE enables the experimenter to code and to capture in real time data linked to user's activity (for example, speech, gestures). Activity is defined as events (single event, like coughing for 2 sec) or durations (with a start and a stop event, like consulting an online help for 30 sec) which are assigned to keys of the keyboard of the server. During the test, the experiment hits the keys he has previously assigned to events or durations which results in marks that are synchronized with other data.

### Physiological data

AMUSE enables the experimenter to continuously measure physiological signals. Those signals give various information on user's emotional state, allowing to measure arousal and valence of felt emotions when used and analyzed appropriately [10]. AMUSE integrates the ProComp system from Thought Technology Ltd [11]. Up to eight sensors can be used: electromyography (EMG), electroencephalography (EEG), electrocardiography (ECG), blood volume pulse, heart rate, respiration, skin temperature and skin conductance (SC). The sensor information is digitally sampled. The resulting information is sent to the computer via a fibber-optic cable.

Moreover, AMUSE enables the experimenter to record a digital video of the test progress and to synchronize it with all the other recorded data.

## Modes and functionalities

AMUSE functions through three different modes. The first one is dedicated to the preparation of the test. In this mode, the experimenter defines the equipments to be used (eye-tracker, mouse and keyboard tracker, electro-physiologic sensors) and the events and durations to be recorded. The second mode concerns the running of the test and the recording of the data. In this mode, the user is asked to achieve a given scenario, like in a classic ergonomic user test. All recorded data are saved to an Excel® file for further statistical analysis. Figure 1 gives an illustration of such a test where the electro-physiologic sensors used were the skin conductance (left hand) and the electromyography (right hand) ones. The eye-tracker video camera can be seen below the PC's screen. The third mode permits to replay the collected data after the test.



Figure 1. A user testing a web site with AMUSE.

Screens that have been seen by the user, his/her eye gaze, his/her physiological data, the moves of the mouse, keys struck on the keyboard, events and durations entered by the experimenter during the test can be visualized at the same time. Combined with the synchronized video recording of the test, this replay enables the experimenter to achieve precise and exhaustive analyses of collected data.

Figure 2 is a sample screenshot of the replay mode of AMUSE. First of all, the time progress of the replay mode can be seen at the bottom of the screenshot (in the rectangle) or by following the progression of the grey bar from left to right in frame 4 (see arrow). The upper frame (frame 1) is a copy of the screen seen by the user at any instant of the test. Inside this frame, two icons appear (in the circle): one eye and one cursor, which symbolize user's eye gaze and mouse cursor during the test. In the replay mode, these two icons exactly reproduce the user's eye and mouse movements as recorded during the test. Each new window seen by the user during the test is immediately refreshed in frame 4 at the equivalent instant in the replay mode. Frame 2 and frame 4 show the markers of events and durations inserted by the experiment during the test. In frame 2, they appear as

textual (which event/duration) and time (when) information whereas in frame 4, they appear as graphic information (vertical bars for events and horizontal bars for durations). Additionally, struck keys of the keyboard appear on frame 2. Frame 3 shows the recorded electro-physiologic signals. In this test, electromyography of the right hand (upper curve) and skin conductance (lower curve) signals were recorded.

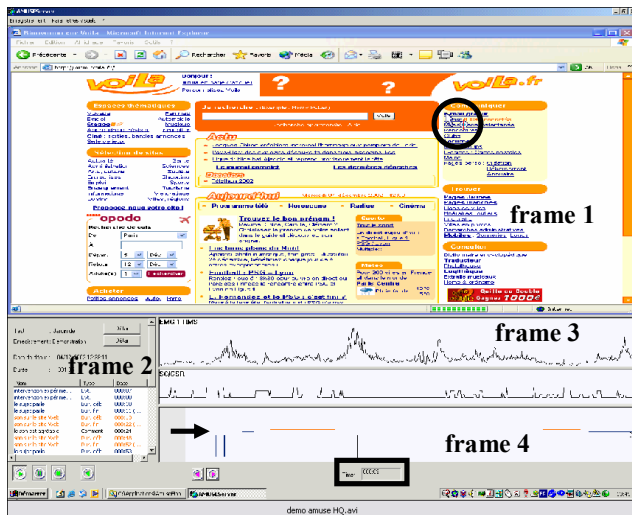


Figure 2. A screenshot of the replay mode.

The combination of the information on these four frames allows the experimenter to precisely analyze any sequence of the test. For example, user's eye gazes can be jointly analyzed with his/her skin conductance, in order to explore whether looking at some specific information, graphics or images increases his/her stress.

### AN APPLICATION: TEST OF ECAS

Considering the ongoing progresses of natural-language dialogue technology and of virtual characters representation, several researchers stand up for the fact that ECAs represent a promising mean for proposing easier and more natural human-machine interactions (e.g., [2]). Additionally, interaction with ECAs appears to be a paradigm where emotion recognition and expression could be most logically studied since it represents one of the most anthropomorphic man-machine interactions up to date [1].

One of the parameters controlling the naturalness and the expressivity of ECAs is synthetic speech. Although the variability of associations between faces and synthetic voices is known to influence users' perception of the quality and coherence of talking faces [9], nothing is known on the influence of synthetic speech on users' emotional reactions to ECAs. In order to illustrate some of the possibilities of AMUSE, we will here present briefly an experiment using this tool and aiming at studying the emotional reaction of adults and teenagers to an ECA within the context of a theatre play game.

### Methodology

The game proposed to users introduces a virtual theater partner (the simulated ECA) which can play opposite to the user who is asked to learn a scene of the play "Le Petit Prince", from Antoine de St Exupery [12]. The interaction uses the Wizard of Oz methodology, in the sense that users think they are really interacting by voice with a virtual character, whereas the reactions of the character are in fact controlled in real time by a hidden experimenter.

The test consisted in three sessions of approximately 2 minutes each where each user had to play the character of "the little prince", while the character of "the business man" was played by the ECA. In each session, the ECA had a different voice: two synthetic ones of different qualities (high and low quality), and a natural one. Audiovisual sequences were pre-treated (for speech-lip movements synchrony) prior to the test and stored on a PC. During a given session, the user had to play his part and the ECA automatically played his own at the correct rhythm as if a speech recognition and a dialogic model were used. Figure 3 gives a screenshot of what users saw on the PC's screen during the test. Each of the three paragraphs consisted in one of the three sessions.

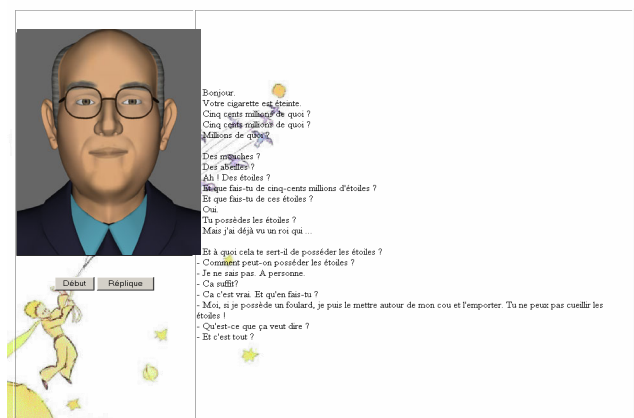


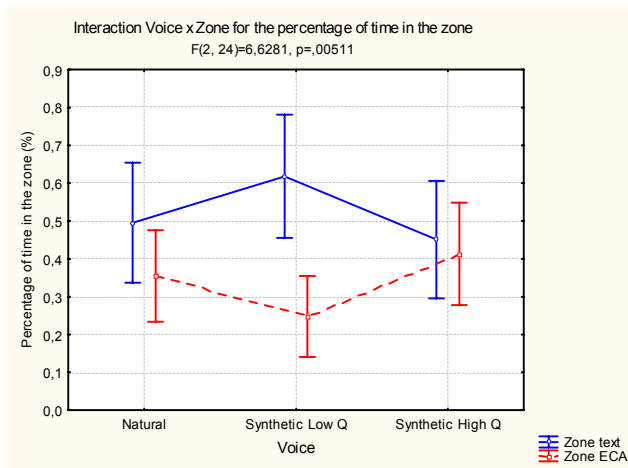
Figure 3. A screenshot of the screen seen by user during the test.

During this test, AMUSE was used to record user's skin conductance and eye gaze. After each session, users were asked to describe the emotional state they experienced by interacting with the ECA by using the PrEmo character grid [3]. Additionally, audiovisual recordings of users were made during the tests in order to achieve further analyzes of facial mimics and voice intonations during their interactions with the ECA. Nine adults and four teenagers participated to this test.

### Some results

Considering that this experiment is only presented as an example of use of AMUSE, only a few results illustrating the possibilities of AMUSE will be presented here. Figure 4 shows the percentage of time of the eye gaze in the two zones of interest of the screen (zone of the ECA and zone of

text) for each condition: for natural voice, synthetic voice of low quality and synthetic voice of high quality. A significant interaction Voice x Zone can be observed ( $F(2,24)=6.28$ ,  $p<0.01$ ), showing that users spent significantly less time looking at the ECA when his voice was a synthetic low-quality one. It seems that users more focused their attention on the text when the ECA was of low quality, therefore being poorly attractive.



**Figure 4. Percentage of time of the eye gaze in the two zones of the screen.**

By crossing eye gaze and electro-physiologic data, AMUSE allows to analyse the values of the mean of the skin conductance signals when the eye gaze are in a given zone of the screen. Such an analysis was conducted for the text and the ECA zones of the screen, for the three voice conditions, and showed that there were no significant differences between zones and between voices ( $F(2,24)=0.99$   $p=0.39$ ). Therefore, although voice quality seems to have an effect on users' eye gaze (driving user's attention on the screen), it seems to have no effect on their emotional reaction as measured by skin conductance. The analysis of the PreMo answers confirmed this finding.

## CONCLUSION

AMUSE is an innovative tool that enables experimenters to gather subjective and objective data about the behavior of a user. Beyond the help it brings to conduct classic ergonomic user tests of Computer-Human Interfaces, it allows to evaluate their affective dimension, notably by analyzing in parallel several sources of information, such as eye gaze and electro-physiologic signals. However, AMUSE is focused on the instantaneous reaction and behavior of users during tests and its use has to be completed by *a posteriori* evaluations such as self-assessments methods, questionnaires or interviews in order to get the full picture of the ergonomic and affective evaluation of new interfaces.

## ACKNOWLEDGMENTS

We wish to thank Pierre Davy for his technical support in the design of this tool, Gaspard Breton, Danièle Pelé and Paul Bagshaw for providing the talking head, Valérie Maffiolo and Nathalie Pican for their help in the design and the running of the test.

## REFERENCES

1. G. Ball and J. Breese. Emotion and personality in a conversational agent. In S. Prevost J. Cassell, J. Sullivan and E. Churchill, editors, *Embodied Conversational Characters*. MITpress, Cambridge, MA, 2000.
2. Bickmore, T. W., Unspoken Rules of Spoken Interaction, in: *Communications of the ACM* 47 (2004) 4, 38-44.
3. Desmet, P.M.A., Hekkert, P., Jacobs, J.J. When a car makes you smile: Development and application of an instrument to measure product emotions. In: S.J. Hoch, R.J. Meyer (Ed.), *Advances in Consumer Research*, 27 (2000), 111-117.
4. Hedonics-affective human factors design, Special Issue in *Ergonomics*, Vol. 46, 13-14, Martin G. Helander and Ming Po Tham (Eds.), 2003.
5. Hugdahl, K. and Stormark, K.M. Emotion Modulation of Selective Attention: Behavioral and Psychophysiological Measures. In *Handbook of Affective Sciences*, Oxford University Press, 2003.
6. Jordan, P. *Designing Pleasurable Products: An Introduction to the New Human Factors*. London: Taylor & Francis, 2000.
7. Lang P. *The cognitive Psychophysiology of emotion: anxiety and the anxiety disorders*. Hillsdale NJ: Lawrence Erlbaum, 1985.
8. Mano H. Assessing emotional reactions to TV ads: a replication and extension with a brief adjective checklist. *Advances in consumer research*, 23 (1996), 63-69.
9. Mersiolo, M., Chateau, N. & Maffiolo, V. Talking Heads: Which Matching between Faces and Synthetic Voices? *Proc. Fourth IEEE International Conference on Multimodal Interfaces (ICMI'2002)*, 2002, 69-74.
10. Picard, R.W. *Affective Computing*. MIT Press, 1997. see Picard, 1997, pp. 141-163
11. Thought Technology Ltd. <http://www.thoughttechnology.com/>
12. de Saint Exupéry, A. *Le Petit Prince*, Paris: Reynal & Hitchcock, 1943.
13. Vick, R. M. and Ikehara, C. S. Methodological Issues of Real Time Data Acquisition from Multiple Sources of Physiological Data. *Proc. 36<sup>th</sup> Hawaii International Conference on System Sciences*, 2003.