Double level analysis of the Multimodal Expressions of Emotions in Human-Machine Interaction

Jean-Marc Colletta [1], Ramona Kunene [1], Aurélie Venouil [1], Anna Tcherkassof [2]

[1]Lidilem, Université Stendhal, BP25, 38040 Grenoble Cedex 9, France
[2] LPS, Université Pierre Mendès France, BP47, 38040 Grenoble Cedex 9, France
e-mail: Jean-marc.colletta@u-grenoble3.fr, kuneneramona@yahoo.com, a.venouil@free.fr, anna.tcherkassof@upmf-grenoble.fr

Abstract

This paper presents the method and tools applied to the annotation of a corpus of multimodal spontaneous expressions of emotions, aimed at improving the detection and characterisation of emotions and mental states in human-machine interaction. The annotation of multimodal corpora remains a complex science as the preparation of the analysis tools have to be in line with the objectives and goals of the research. In human expressions and emotions the verbal and non verbal behaviour all play a crucial role to reveal the mental state of a speaker and as such voice, silences, hesitations from the verbal aspect, and every movement from the scratching of one's eye to the movement of toes from the non verbal aspect, have to be taken into consideration. The physical description of the bodily movements, although necessary, remains approximative when based on 2D and lacks the analytical aspects of human behaviour. In this paper we define a two-level procedure for the annotation of the bodily expressions of emotions and mental states, as well as our annotation grid for speech cues and body movements.

1. Introduction

This paper presents the annotation procedure of a corpus of multimodal spontaneous expressions of emotions and mental states in human-machine interaction.

The corpus collection was part of a study on the fusion of multimodal information (verbal, prosodic, facial, gesture, posture and physiology) to improve the detection and characterisation of expressions of emotions in human-machine interaction (Le Chenadec, Maffiolo & Chateau, 2007).

The overall objective was to develop computer systems which can « perceive and understand human behaviour and respond appropriately » (Le Chenadec, Maffiolo, Chateau & Colletta, 2006). In the optic to develop these affective computer systems which detect and characterise expressions of emotions and mental states, the data collected had to reflect the multimodal character of human behaviour.

The annotation considerations were to identify the mental and emotional states from a video corpus of 18 adults.

The next section presents an overall view of the data collection and the methodological aspects of this study. The following sections discuss annotation procedures and present the annotation scheme we created for this study.

2. Elicitation Methodology

The experimental setup has been well presented in Le Chenadec, Maffiolo, Chateau & Colletta, (2006) and Le Chenadec, Maffiolo & Château, (2007). Here we give a brief recapitulative.

The objectives of the corpus collection were threefold: the range of emotional and mental states had to be widest as possible, emotions and mental states had to be expressed freely and spontaneously, and expressions had to be multimodal through vocal, gesture, postural, facial, physiological behaviour.

This experiment was conducted in a laboratory test platform based on the Wizard-of-Oz methodology, in which an interaction between a human and a virtual character could be set up. In this experiment the interaction was on the repetition of a play. The instructions given to the users were to play three scenes of *Don Quixote de la Mancha,* written by M. de Cervantes (1605). The human subject was to play the part of Sancho Panza and had to give his cue to the virtual character as Don Quixote. Cues from the virtual actor were controlled by the experimenter in real time. The experimenter simulated an autonomous system.

In order to elicit spontaneous emotional expressions of users, different system bugs were designed: uncoordinated movements or stammering of the virtual actor and the request to the user was to repeat his cue, or if the system displayed "lost data", the request to the user was to repeat one of the three scenes of the act. From the users' perspective, some bugs were clearly related to a system's failure, other bugs were perceived to be a result of their mistakes of their cues. They were expected to express the emotional feelings and mental states they experienced as a result of being confronted with these bugs, which were designed to be funny, or boring, repetitive or deeply annoying.

The multimodal behaviour of each user was recorded with two digital cameras (head-only and upper body) and a microphone. Eighteen actors (nine females and nine males, aging from 25 to 50 years) took part in the experiment. Interview recordings lasted 1h15mn for each participant.

The data collected during the experiment were completed by the gathering of the user's viewpoint immediately after the interaction with the virtual character (Le Chenadec, Maffiolo & Chateau, 2007). Each user was asked to comment what he/she felt during the interaction while viewing its recording, and to determine the starting and ending time where he/she experienced these feelings. A subsequent interview was conducted with a close relative of each user. The recording of the interaction was played back to this relative who was asked to comment on the behaviour of the user using the same method.

Finally, a categorisation experiment was conducted at the LPS laboratory, Université Pierre Mendès France, Grenoble. The same recordings were played back twice to 90 third-party observers, all students in social psychology. The first viewing allowed each subject to familiarise him/herself to the idiosyncratic behavioural characteristics of the user observed. During the second viewing, he/she was asked to stop the video each time he/she observed that the user felt something, i.e. experienced an emotional or a cognitive state. He/she then had to attribute an emotional or cognitive value to the observed behaviour and indicate his/ her starting and ending time.

The next section discusses the key factors applied to the annotation process of the data collected during the experiment.

3. Transcription Considerations

Currently, several researchers are interested in the multimodal complexity processes of oral communication. This issue has brought about increased interest to researchers aiming to transcribe and annotate different kind of multimodal corpora. Some researchers, as Abrilian (2005), work on the annotation of emotional corpora in order to examine the relation between multimodal behaviour and natural emotions. Other researchers working in the field of autism (*inter alia* Grynszpan, Martin & Oudin, 2003) or language development (Colletta, 2004; Colletta et al, this symposium) also takes into consideration these multimodal clues in their studies. Researchers in computer sciences take into account the multimodal clues in order to improve the ECAs – Embodied Conversational Agents – (Hartmann, Mancini & Pelachaud, 2002, 2005; Ech Chafai, Pelachaud & Pelé, 2006; Kipp, Neff & Albrecht, 2006; Kipp et al., 2007; Kopp et al., 2007; Vilhjalmsson et al., 2007).

It is without doubt that these methods and tools of annotation have paved the way for more interesting exploratory means to study multimodal corpora in detail. However, some theoretical and methodological difficulties still arise when one tries to annotate body movements. We will discuss these points in the following section 3.2.

The 18 recordings of the interactions between the subjects and the theatrical application, treated by the Lidilem laboratory, Université Stendhal, Grenoble, were specifically dedicated to the obtaining of the multimodal expressions of spontaneous emotional and mental states.

Two kinds of annotations were conducted: an annotation of each user's speech as well as other paraverbal phenomena – prosodic and voice considerations –, and an annotation of their corporal behaviour throughout the repetition of the play experiment.

3.1. Verbal and prosodic annotation

Linguistic and prosodic attributes often betray the emotional as well as the mental state of the speaker's mind. Each emotion has its words and verbal expressions, as research on the semantics of emotion show (Galati & Sini, 2000; Plantin, 2003; Tutin, Novakova, Grossmann & Cavalla, 2006). Stronger cues are supported by the voice features (Lacheret-Dujour & Beaugendre, 1999; Aubergé & Lemaître, 2000; Scherer, Bänziger & Greandjean, 2003; Keller et al., 2003; Shochi, Aubergé & Rilliard, 2006). In fact, all aspects of prosody may contribute to express emotional and mental states: pitch, intensity, speech rate, hesitations, grunts, various mouth and throat noises, etc.

Our verbal annotation was done on the software *PRAAT* developed by P. Boersma and D. Weenink [1]. As Table 1 shows (see: annexures), we did not code for pitch, intensity or rate as this data can be directly collected from the speech signal analysis. We coded for silent and filled pauses, linguistic errors, unexpected articulation of words, false starts, repetitions, laughs, coughs and sighs, all linguistic or posodic cues which may be an indication of reflection, embarrassment or various emotions.

3.2. Non-verbal annotation

The verbal transcriptions were aligned and imported to the software *ANVIL* developed by M. Kipp [2]. All recordings with their corresponding visual components were annotated accordingly in respect of the non-verbal performance of the subject.

As gesture researchers have already demonstrated in the past, all bodily movements may help express attitudes, emotional and mental states (Feyereisen & De Lannoy, 1985; Kendon, 1990, 2004; Feldman & Rimé, 1991; Descamps, 1993; Cosnier, 1994; Plantin, Doury & Traverso, 2000; Knapp & Hall, 2002). Attitudes and postures were correlated to mental states, pathologies and emotional disposition of the subjects;

[1] Available from http://www.fon.hum.uva.nl/praat/
[2] Available from http://www.anvil-software.de/

the gaze contributes to the expression of emotion and its appearance was correlated to the levels of activation or attention of the subject; the facial expressions exteriorise the whole range of emotions and feelings. Finally, among gestures, we can observe that some appear more frequently in stressful situations and are correlated to anxious states and certain affects: the gestures which are self centred, which include the gestures of self-contact (rub oneself on the face, scratching oneself, massaging oneself) and the gestures of manipulation of objects (playing with his/her keys, fiddling with his/her pen).

To annotate the non-verbal behavioural features of the subjects in this study, the coding scheme used was divided in 16 tracks (see Figure 1 and Table 2: annexures) representing all different parts of the human body. Our annotation grid was thus split into:

(i) self-contact gestures and auto-manipulations;
(ii) posture attitudes and changes (2 tracks);
(iii) head gestures (2 tracks);
(iv) gaze direction and changes (2 tracks);
(v) facial expressions (2 tracks);
(vi) torso movements;
(vii) shoulders movements;
(viii) arms location and movements;
(ix) hand gestures (2 tracks);
(x) lower body movements;
(xi) gestures performed by the actor while giving his clues to the animated character and part of his acting.

Each subject file had two subfiles; a video with both the face and body which allowed to annotate all the above mentioned body part, and a purely facial video to allow for precise, accurate and detailed coding of facial expressions.

From an etymological perspective (Pike, 1967), to obtain an annotation of the mental and emotional state behaviour of the speaker, an *etic* approach is necessary which will emphasise the physical aspects of the movement and allow for a microanalytical description. Researchers in gesture synthesis all agree on the necessity to rely on physical and accurate descriptions of the body movements. The transcription tools they propose all annotate for the body parts, as: gesture, gaze, head, torso, face, legs, lips and other behaviour (Vilhjalmsson et al., 2007). They also annotate for various location and movement parameters. For instance, to annotate for gesture expressivity, Hartmann, Mancini and Pelachaud (2005) distinguish between overall activation, spatial and temporal extent of the movement, fluidity (smooth *vs*. jerky), power (weak *vs*. strong) and repetition. To annotate for hand gestures, the researchers trying to unify a multimodal behaviour generation framework called the "Behavior Markup Language" (Kopp et al., 2006; Vilhjalmsson et al., 2007) mention the following parameters: wrist location, trajectory of movement, hand shape, hand orientation. Kipp, Neff & Albrecht (2006) propose to annotate for "handedness", trajectory, hand location (height, distance, and radial orientation), position of the arm (arm swivel) and hand to hand distance for a two hands gesture. When the annotation of hand gestures aims at studying the relationship between gesture and speech (see McNeill, 1992, 2005; Colletta, 2004), it also requires temporal information about the phases of the gesture phrase realisation, as first described by Kendon (1972, 1980) and integrated in gesture synthesis by Kipp, Neff & Albrecht (2006).

In our grid (Table 2) , the *etic* approach is displayed under all tracks except those which are subtitled "function", and it gives information on :

(i) the body part and its location (for an arm or a hand gesture),
(ii) direction of the movement,
(iii) characteristic of the movement (swaying, frowning, shrugging, etc.),
(iv) shape of the movement (for a hand gesture),
(v) speed of the movement, and
(vi) frequency of occurrence when the movement is repeated.

However the *etic* approach is not sufficient to present a comprehensive description of bodily behaviour. Kendon (1990), in line with other researchers, have pointed that in everyday life we "read" bodily behaviour of others through mentally precategorised concepts; such as laughing, smiling, ease, nodding, pointing, gesturing, miming, etc.

Each concept covers a range of behaviours, whether small or large, whose physical properties may vary in proportion. For instance, I can smile with a closed mouth or with an open mouth; I can express a subtle smile or a broad smile; I can express a mouth-only smile or be all smiles, etc. Yet all these various forms of smiles are examples of the same broad expressive category called "smile". As for a pointing gesture, I can point with a hand or a head or the chin; I can point to an object or a person present in the physical setting, or to a direction; I can point to an object or person with an extended hand or just with an extended index finger; I can point once to an object or person, or point repetitively to it, etc. There again, all these various forms of pointing share the same function and are exemplars of the category called "pointing gesture".

At this point, it is worth noting that the researchers who currently aim at unifying a multimodal behaviour generation framework for ECAs (Vilhjalmsson et al., 2007) « have proposed knowledge structures that describe the form and generation of multimodal communicative behaviour at different levels of abstraction ». The first level represents the interface between planning communicative intent and planning the multimodal realisation of this intent, and is mediated by the "Functional Markup Language" (FML). The second level represents the interface between planning the multimodal realisation of a communicative intent and the realisation of the planned behaviours, and is mediated by the "Behaviour Markup Language" (BML). Although the FML remains largely

undefined in the authors work, the FML/BML distinction surprisingly resembles Kenneth Pike's distinction between the *emic/etic* levels of behaviour description.

In our view, a more *emic* approach (Pike, 1967) is thus essential to annotate the body movements that express the mental and emotional state behaviour of the speaker, and to complement the *etic* physical description of these movements. In our grid (see Table 2: annexures), this approach is displayed under all tracks which are subtitled "function" and it serves as an indication of:

(i) a general behaviour or attitude (scratching, touching, handling, comfort posture…);
(ii) a significant head movement (head nod, head shake, head beat, deictic or pointing movement);
(iii) a gaze behaviour (waiting, reading, staring, scanning);
(iv) a significant facial expression (smile, laughter, biting, pursing, licking lips, pouting);
(v) a coverbal hand gesture (deictic or pointing movement, beat, iconic gesture, metaphoric gesture, interactive gesture.).

During the annotation process, every body movement was annotated for its *etic* or physical properties as well as for its *emic* properties or emotional/function properties.

4. Transcription and Validation

Coders selected for the annotation had previous experience in gesture and emotion studies. Additional training on annotation tool was included to familiarise them with the *ANVIL* software as well as with the video data. File sequences were initially transcribed manually on *Excel*, in which the coders would first examine the video files and have a global view of the frequency and nature of movements in order to prepare the relevant grid.

The non verbal transcription was then carried out in parallel by two coders. Each coder annotated independently from the other coder. In most cases, the validation of an annotation scheme is based on the comparison of the annotations done by two independent coders. This method is useful to test the validity of an annotation scheme, but it does not allow to check and to stabilise the analysis of a corpus at the end of an annotation procedure. Indeed, in our case, it is not a question of testing a body movement annotation grid, but it is rather a question of validating the annotation of a multimodal corpus before using the results of the annotation in a study on the fusion of multimodal information (Le Chenadec, Maffiolo & Château, 2007). As a consequence, a third coder was asked to finalise the annotation from choices made by both coders and decide in case of disagreement.

Having a two-stage process with the independent coding as well as the decision stage cannot ensure that this analysis procedure is a hundred percent conclusive. To annotate for emotions and mental states is to observe the whole body, including the problem of identifying the movements, which does not arise when we annotate for precise gestures (e.g., the coverbal hand gestures). On the other hand, another means of validation is to cross-check the information resulting from the annotation by the coders with other data sources. For this study on the fusion of multimodal information, the other available data source is (1) the collection of the user's viewpoint after the experiment, completed by interviews with their relatives, and (2) a categorisation experiment conducted with 90 third-party observers (see section 2 for more details). In the end, it will be most interesting to compare the transcriptions by the three coders to the analysis performed by the users and their relatives on one side, and by the 90 students, on the other side.

5. Final remarks

Our paper describes the method and the analysis tools applied as well as the annotating considerations we employed. Our aim is to enhance the understanding of the technical issues surrounding the annotation of a multimodal corpus. Annotating mental and emotional states of mind in adults requires a vigorous approach and attention to detail. The objectives of this research required the minute examination of: the voice, linguistic features, sounds or the absence of sounds as all these play a role in revealing the emotional and state of a speaker. In verbal annotation, we observed all the linguistic and prosodic cues as they offer us a window to the state of nervousness, anxiety, irritation, humour, etc.

The non verbal annotation also required a vigorous if not somewhat lengthy approach. If one seeks the understanding of gesture related to speech it would be much simpler to annotate for hand and head movements, and stick to communicative or representational gesture. In this study, the quest for emotions and human mental states showed that each and every part of the body from the head to the toes has a story to reveal. The grid used on *ANVIL* enabled us to annotate this rather complex set of movements as the human speaker is in constant motion, from scratching his head in anxiety to smiling in contentment.

Our analysis procedure aimed at using the double level (*etic/emic*) annotation, which we hope, will help to enhance in the designing of annotation tools. The missing puzzle remains in the cross-validation from several data sources.

6. Acknowledgements

This research was conducted and financed by the France Telecom R & D, Lannion. The authors thank Valérie Maffiolo and Gilles Le Chenadec for the designing of the experiment and for contributing to the creation of the annotation grid.

7. References

Abrilian, S. (2005). Annotation de corpus d'interviews télévisées pour la modélisation de relation entre comportements multimodaux et émotions naturelles. *6ème colloque des jeunes chercheurs en Sciences Cognitives* (*CJCSC'2005), Bordeaux, France.*

Aubergé, V., Lemaître, L. (2000). The Prosody of Smile. In *Proceedings of the ISCA Workshop on Speech and Emotion, Newcastle, Northern Ireland, sept. 5th-7th, 2000*, pp. 122--126.

Colletta, J.-M. (2004). Le développement de la parole chez l'enfant âgé de 6 à 11 ans. Corps, langage et cognition. Hayen, Mardaga.

Cosnier, J. (1994). *Psychologie des émotions et des sentiments*. Paris, Retz.

Descamps, M.-A. (1993). *Le langage du corps et la communication corporelle*. Paris, P.U.F.

Ech Chafai, N., Pelachaud, C., Pelé, D. (2006), Analysis of Gesture Expressivity Modulations from Cartoons Animations. In *LREC 2006 Workshop on "Multimodal Corpora", Genova, Italy, 27 May*.

Feldman, R., Rimé, B., Dir. (1991). *Fundamentals of non verbal behaviour*. Cambridge, Cambridge University Press.

Feyereisen, P., De Lannoy, J.-D. (1985). *Psychologie du geste*. Bruxelles, Pierre Mardaga.

Galati, D., Sini, B. (2000). Les structures sémantiques du lexique français des émotions. In. C. Plantin, M. Doury, V. Traverso, *Les émotions dans les interactions*. Presses Universitaires de Lyon.

Grynszpan, O., Martin, J.C., Oudin, N. (2003). On the annotation of gestures in multimodal autistic behaviour. In *Gesture Workshop 2003*, Genova, Italy, 15-17 April.

Hartmann, B., Mancini, M., Pelachaud, C. (2002). Formational Parameters and Adaptative Prototype Instanciation for MPEG-4 Compliant Gesture Synthesis. *Computer Animation Proceedings, Genève, June 2002*.

Hartmann, B., Mancini, M., Pelachaud, C. (2005). Implementing Expressive Gesture Synthesis for Embodied Conversational Agents. *Gesture Workshop, LNAI, Springer, May 2005*.

Keller, E., Bailly, G., Monaghan, A., Terken, J., Huckvale, M. (2003). *Improvements in speech synthesis*. Chichester, UK, John Wiley.

Kendon, A. (1972). Some relationships between body motion and speech. In A.W. Siegman et B. Pope (eds.), *Studies in dyadic communication*. Elmsford, NY, Pergamon Press, pp. 177--210.

Kendon, A. (1980). Gesticulation and speech, two aspects of the process of utterance. In M.R. Key (ed.), *The relationship of verbal and nonverbal communication*. The Hague, Mouton, pp. 207--227.

Kendon, A. (1990). *Conducting interaction. Patterns of behavior in focused encounters*. Cambridge, Cambridge University Press.

Kendon, A. (2004). *Gesture. Visible action as utterance*. Cambridge. Cambridge University Press.

Kipp, M., Neff, M., Albrecht, I. (2006). An Annotation Scheme for Conversational Gestures: How to economically capture timing and form. In *Proceedings of the Workshop on Multimodal Corpora (LREC'06)*, pp. 24--27.

Kipp, M., Neff, M., Kipp, K.H., Albrecht, I. (2007). Towards Natural Gesture Synthesis: Evaluating gesture units in a data-driven approach to gesture synthesis. In C. Pelachaud et al. (eds.), *Intelligent Virtual Agents 2007, Lecture Notes in Artificial Intelligence 4722*. Berlin, Springer-Verlag, pp. 15--28.

Knapp, M**.**, Hall, J. (2002). *Nonverbal communication in human interaction*. Harcourt College Publishers.

Kopp, S., Krenn, B., Marsella, S., Marshall, A.N., Pelachaud, C., Pirker, H., Thórisson, K.R., Vilhjálmsson, H. (2007). Towards a Common Framework for Multimodal Generation: The Behavior Markup Language. In J. Gratch et al. (eds.), *Intelligent Virtual Agents 2006, Lecture Notes in Artificial Intelligence 4133*. Berlin, Springer-Verlag, pp. 205--217.

Lacheret-Dujour, A., Beaugendre, F. (1999). *La prosodie du français*. Paris, CNRS Editions.

Le Chenadec., G., Maffiolo V., Chateau N. (2007). Analysis of the multimodal behavior of users in HCI : the expert viewpoint of close relations. *4th Joint Workshop on Multimodal Interaction and Related Machine Learning Algorithms, 28-30th June, Brno, Czech Republic.*

Le Chenadec, G.., Maffiolo, V., Chateau, N., Colletta, J.M. (2006). Creation of a Corpus of Multimodal Spontaneous Expressions of Emotions in Human-Interaction. In *LREC 2006, Genoa, Italy.*

McNeill, D. (1992). Hand and mind. What gestures reveal about thought. Chicago, University of Chicago Press.

McNeill, D. (2005). *Gesture and thought.* Chicago, University of Chicago Press.

Pike, K.L. (1967). *Language in relation to a unified theory of the structure of human behavior*. Janua Linguarum, series maior, 24. The Hague: Mouton.

Plantin, C. (2003). Structures verbales de l'émotion parlée et de la parole émue. In J.-M. Colletta, A. Tcherkassof, *Les émotions. Cognition, langage et développement*. Hayen, Mardaga, pp. 97--130.

Scherer, K.R., Bänziger, T. Grandjean, D. (2003). L'étude de l'expression vocale des émotions : mise en évidence de la dynamique des processus affectifs. In J.M. Colletta, A. Tcherkassof, *Les émotions. Cognition, langage et développement*. Hayen, Mardaga, pp. 39--58.

Shochi, T., Aubergé, V., Rilliard, A. (2006). How Prosodic Attitudes can be False Friends: Japanese vs. French social affects. *Proceedings of Speech Prosody 2006, Dresden*, pp. 692--696.

Tutin, A., Novakova, I., Grossmann, F., Cavalla, C. (2006). Esquisse de typologie des noms d'affect à partir de leurs propriétés combinatoires. *Langue Française*, 150, pp. 32--49.

Vilhjalmsson, H., Cantelmo, N., Cassell, J., Chafai, E.N., Kipp, M., Kopp, S., Mancini, M., Marsella, S., Marshall, A.N., Pelachaud, C., Ruttkay, Z., Thórisson, K.R., van Welbergen, H., van der Werf, R.J. (2007). The Behavior Markup Language: Recent Developments and Challenges, in C. Pelachaud et al. (eds.), *Intelligent Virtual Agents 2007, Lecture Notes in Artificial Intelligence 4722*, Berlin, Springer-Verlag, pp. 99--111.

Annexures

Type of annotation	**Name of phenomenon**	**Definition**
Prosody	Silent pause	pause in the middle of a speech segment
	Intelligible pause	silence voluntarily added in the middle of a speech segment
	Pause filler	"euh" ou "hum"
Linguistic	Commentary	commentary on the interaction
	Error	error in syllable or word pronounciation
	Unexpected articulation	pronounciation of an unsual final syllable with a silent "e.
	False start	a "*" attached to the word + annotate the complete word sequence
	Elision	presence of elision
	Recovery	reformulation of a portion of a speech segment
	Repetition	repetition of a portion of a speech segment
	Incomprehensible words	transcription of an impossible word or speech segment
Dialogue	Repétition	repetition of the identical
	Reformulation	repetition of response with other terms
Sounds	Sounds of the system	
	Speech cuts	the virtual actor cuts the live actor's speech
	cough, throat, mouth	cough, throat clearing, noise made by the mouth
	Laugh	
	Exhalation, breath, sigh	
	Inhalation	

Table 1 : Verbal and prosodic annotation grid

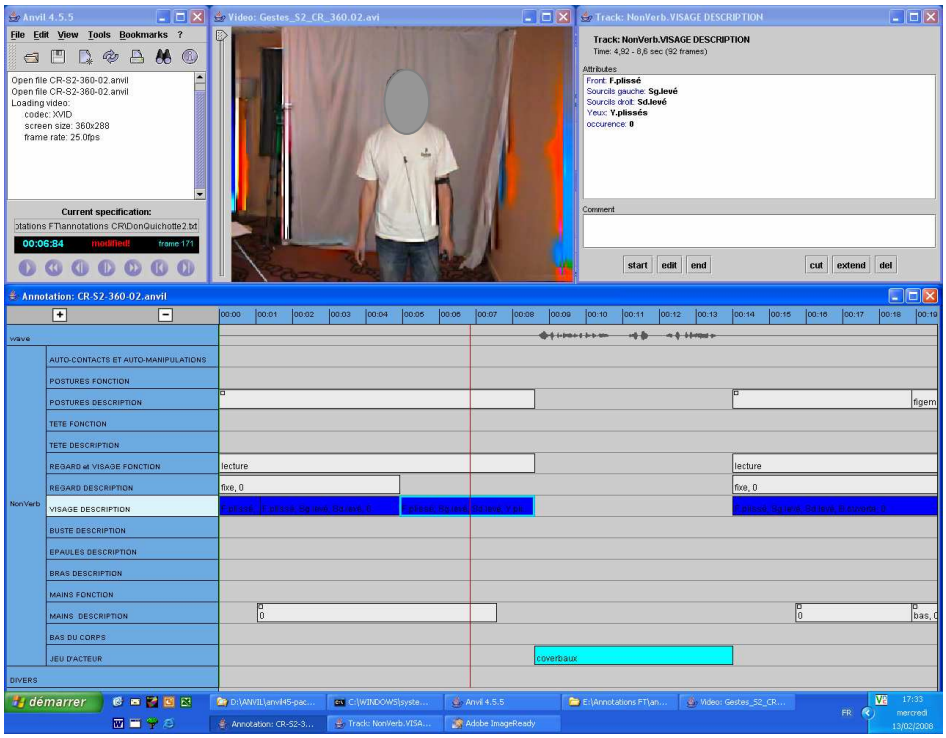

Figure 1: Anvil Screen

Type of annotation	Name of phenomenon
1- Self-contact gestures & auto-manipulations	Action: scratch/ touch/ twist/ handle
	Body part location: hair/temple/brow/glasses/ nose
2- Posture (function)	Comfort/ stretching
3- Posture (description)	Pattern : swaying/ complex movement/ freezing
	Leg movements: frontward/ backwards/ left/ right
	Speed: slow/ normal/ fast
4- Head (function)	Movement : nod/ shake/ beat/ deictic
5- Head (description)	Tilted high/ low
	Turn: left/ right
	Complex movement: front / backward
	Single movement: up / down
	Single movement: front / backward
	Single tilt: left/ right
	Single side-turn: left/right
	Speed: slow/ normal/ fast
6- Gaze (function)	Characterisation: waiting/ reading/ staring/ scanning
7- Gaze (description)	Direction: up/ down
	Direction: left/right
	Movement: sweeping/ rolling eyes
	Speed: slow/ normal/ fast
8- Face (function)	Smile, laughter/ biting/ pursing/ licking lips/pouting
9- Face (description)	Brows: frowning
	Left eyebrow: raising / frowning
	Right eyebrow: raising/ frowning
	Eyes: closing / opening/ wide opening/ rolling/ blinking/ winking
10- Torso (description)	Movement: forward/ backward
	Movement: left/right
	Unsteady movement
	Bend: forward/ backward
	Turn:left/ right
	Twist: left/ right
	Side: left/ right
	Position: protruded/ retracted
	Speed: slow/ normal/ fast
11- Shoulders (description)	Identification: left/ right/both
	Description: shrugging/ sagging
	Number: left/ right/ both
	Occurrence: 0 to 5
	Speed: slow/normal/fast
12- Arms (description)	Left-arm direction: going up/down, moving sideways, forwards, backwards, to the side, up, not moving
	Left-arm position: bent, half-bent, stretched out
	Right-arm direction: going up / down, moving sideways, forwards, backwards, to the side, up, not moving
	Right-arm position: bent, half-bent, stretched out
	Both arms action: crossing
	Occurrence: 0 to 5
	Speed: slow/ normal/ fast
13- Hands (function)	Deictic, beat, iconic, metaphoric, interactive
14- Hands (description)	Left hand action: rotation, opening, closing
	Left-hand direction: up/ down/left/ right/ forward/ backward
	Left-palm direction: Left-hand direction: up/ down/left/ right/ forward/ backward
	Right hand action: rotation, opening, closing
	Right-hand direction: up/ down/left/ right/ forward/ backward
	Right-palm direction: up/ down/left/ right/ forward/ backward
	Occurrence: 0 to 5
	Speed: slow/ normal/ fast
15- Lower body	Free comments
16- Acting	Mime, exaggerated gestures and expressions

Table 2: Coding scheme for the non verbal annotation grid.