SEMAINE

THE SENSITIVE AGENT PROJECT

D7b

Public access to labelled SEMAINE data

Date: 22 December 2009

Dissemination level: Public

ICT project contract no.	211486
Project title	**SEMAINE** **Sustained Emotionally coloured Machine-human Interaction using Nonverbal Expression**
Contractual date of delivery	*31 December 2009*
Actual date of delivery	*22 December 2009*
Deliverable number	D7b
Deliverable title	Public access to labelled SEMAINE data
Type	Demonstrator
Number of pages	14
WP contributing to the deliverable	WP 6, WP 7
Responsible for task	Michel Valstar (michel.valstar@imperial.co.uk) and Gary McKeown (g.mckeown@qub.ac.uk)
Author(s)	Michel Valstar (michel.valstar@imperial.co.uk) and Gary McKeown (g.mckeown@qub.ac.uk)
EC Project Officer	Philippe Gelin

Table of Contents

1 Executive Summary

We have recorded a dataset of emotionally coloured conversations in the context of the SEMAINE project. The database consists of 24 recordings, which are split into 144 sessions. In total 20 participants are included in the database. The database is publicly available through a web-interface that facilitates easy search and download of the data for researchers.

The experiments are recorded by five high-framerate cameras and four high-quality microphones. Great care was taken to synchronise the recordings from the various sensors, achieving a maximum time difference of 25 microseconds between sensors.

Experimental procedure involved an Operator in one room interacting with a User in a different room through a teleprompter set-up that ensured direct eye contact was possible during the interaction. The Operator played a sequence of emotional roles as defined by the four SEMAINE characters, Obadiah, Poppy, Prudence and Spike. The User interacted with these characters in a natural manner with the single constraint that no questions could be asked.

A substantial proportion of the database has been annotated. All of the User clips are annotated by at least one rater and often as many as four raters. These annotations use trace style continuous ratings, along with utilities provided for converting these into discrete categorical variables if desired. The annotations consist of five dimensions for every annotated clip (Valence, Activation, Power, Anticipation/Expectation and Intensity) and an additional selection of four dimensions from a range of 27 further dimensions. Textual transcripts for the sessions are also provided and each clip has had its textual content analysed using the Linguistic Inquiry and Word Count (LIWC) application; outputs for this analysis and norms for spoken language are provided.

2 Description of data and annotations

We have created a dataset consisting of 24 recordings, which are split into 144 sessions. In total 20 participants are included in the database. The database is publicly avaliable from http://www.semaine-db.eu. In the following sections we will describe in detail how the dataset was created and how it can be accessed by the research community.

2.1 Experiment procedure

The experiment was designed as the first of various situations that mimic the Sensitive Artificial Listener (SAL) interaction. This first experimental situation is known as Solid SAL. The setup is designed to capture naturalistic conversation between two interacting participants. One, the Operator, plays the role of each of the SAL characters in turn, while the User interacts with each of these characters in as natural a manner as possible.

Participants are recruited from a population of undergraduate and postgraduate students of Queen's University Belfast. Only a small number are associated with the SEMAINE research team. Recruited participants are given a written participant information sheet and provide written consent. This is followed by a verbal explanation of the goals of the SEMAINE project and an explanation of what they are required to do in the experiment. It is explained what the nature of each character is and that the session is to last about twenty minutes with an approximate interaction time of five minutes per character. The actual character interaction duration varies depending on the interaction. Participants can change character at any time when they get bored, annoyed or feel that there is nothing more to say to the character. The SAL operator can also request to change character if enough time has passed with a character or a conversation has arrived at a natural conclusion.

Participants are then brought to the recording studio, where they sit down on a chair in the user room and put on their head microphone. The Operator takes her/his place in a separate room and the recording starts. The Operator recites a brief introduction script to the SAL system and asks the User which character they wish to speak to, after which the conversational interaction begins. The Operator is required to act each of the four characters in turn. Poppy is cheerful, optimistic and looks on the bright side of life, Spike is aggressive, confrontational and argumentative, Obadiah is gloomy and has a pessimistic outlook, and Prudence is matter-of-fact, taking a practical view on life.

The Operator attempts to play these character styles in as natural a manner as possible without recourse to a script. Pilot attempts using scripts and learned repertoires resulted in conversation that was not naturalistic and so a more free form approach was adopted instructing Operators to play the roles of the characters. Users interact with the characters as naturally as possible. There is a single constraint: they are not permitted to ask questions to the characters. When questions are asked users are reminded by the Operator that they should not ask questions. In some situations the Operators do not comply with these rules and answered questions and incorporate knowledge from the conversation. However the Operators are instructed that the most important aspect of the task is to create a naturalistic conversation; strict adherence to the rules was only secondary to this and therefore transgressions occasionally occur. Once all four of the characters have interacted with the user the Operator brings the recording session to a close.

After the recording, a debriefing session takes place where the User has a chance to ask more about the system and is asked if there was any part of the procedure that made them feel uncomfortable. This ends the recording session.

2.2 Recording setup

The recording setup for the SEMAINE project is located at Queen's University Belfast. It aims to record audio and video of a user and an operator using five cameras and four microphones, completely synchronous. Video is recorded at 50 frames per second at a spatial resolution of 780 x 580 pixels, while audio is recorded at 48 kHz and 24 bits per sample. Both the User as well as the Operator are recorded from the front by both a greyscale camera and a colour camera. In addition, the user is recorded by a greyscale camera positioned on one side of the User to capture a profile view of their face. To record what the user and the operator are saying, two microphones are used for each: the first is placed on the table and the second is worn by the user/operator. This results in a total of four microphones and thus four recorded channels. The wearable microphone is the main source for capturing the speech and other vocalisations made by the User/Operator, while the room microphones can be used to model and thus reduce background noise.

The User and Operator are located in separate rooms. They can hear each other over a set of speakers, which output the audio recorded by the wearable microphone of their conversational partner. They can see each other on teleprompters. The frontal cameras are placed behind the semi-reflecting mirror. This way, the User and Operator have the sensation that they look each other in the eye. This proved to be very important, as initial tests without teleprompters gave the experimenters the feeling that something very important was missing from the conversation.

It is extremely important to make sure that all sensor data is recorded with the highest possible synchronisation. To do so, an elaborate system was developed that uses the trigger of a single camera to accurately control when all cameras capture a frame. The same trigger was recorded as an audio signal. This allowed us to synchronise the audio and the video sensor data with an accuracy of 25 microseconds. The system is described in full in [LichtenauerEtAl2010].

2.3 Annotations

24 recordings were made using the Solid SAL recording setup. These used 4 different operators and 20 different users (those who had been operators occasionally took the user role). For the purposes of annotation and dissemination the video clips are divided into segments based on character interactions, providing a normal division of each recording session into four character interaction sessions. The division locations are recorded as frame numbers in a text file labelled Session (Number) Timings.

Trace style continuous ratings were made on five dimensions for all clips [Douglas-CowieEtAl2007]. These dimensions are Valence, Activation, Power, Anticipation/Expectation and Intensity these dimensions are consistently used throughout the literature justifying their use in all annotated clips [FontaineEtAl2007]. Raters then chose an additional minimum of four optional rating dimensions from another possible 27 dimensions. The additional dimensions are broken into four categories, as shown in Table 1. The Basic Emotion category stems from a similarly old and extensive literature in emotion research. However, as these basic emotions occur only in certain situations they are included in the optional dimensions [Ekman1999]. The Epistemic States category stems from the work of Baron-Cohen and colleagues including only a small subset of a much larger set of epistemic states [Baron-CohenEtAl2004]. The third category was drawn directly from Bales Interaction Process Analysis [Bales1951]. The final category contains four dimensions that aim to highlight areas of the data where the observed effect and corresponding label differ for a variety of reasons that may be important in assessing the validity of an emotional state. Utilities are provided for converting these continuous labels into categorical labels at provided thresholds.

Basic Emotions	Epistemic States	Interaction Process Analysis	Validity
Anger	(Not)Certain	Shows solidarity	Breakdown of engagement
Disgust	(Dis)Agreement	Shows antagonism	Anomalous simulation
Fear	(Un)Interested	Shows tension	Marked sociable concealment
Happiness	(Not) At ease	Releases tension	Marked sociable simulation
Sadness	(Not) Thoughtful	Makes suggestion	
Contempt	(Not) Concentrating	Asks for suggestion	
Amusement		Gives opinion	
		Asks for opinion	
		Gives information	
		Asks for information	

Table 1: *Dimensional annotation categories*

Transcripts are provided for each of the clips containing text of the spoken interaction and a minimal amount of non-verbal descriptions. Linguistic analysis of these texts is conducted using the Linguistic Inquiry and Word Count (LIWC) program providing an additional 68 linguistic dimensions for the Global conversation, User speech and Operator speech for each clip. Norms are provided for normal spoken language.

3 Web-Accessible Database Interface

3.1 Accessing the database

The database is available from the url http://www.semaine-db.eu. A prospective user would navigate to this page, and would be able to read some information about the database (see Fig. 1). If the user is interested, he or she would proceed by following the 'Request an account' link.

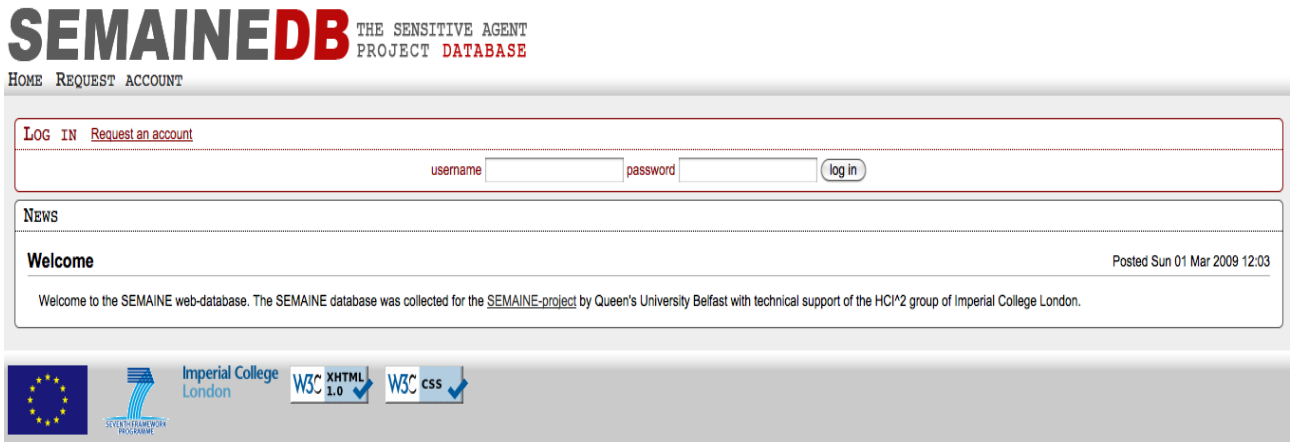

Figure 1: Welcome page of the SEMAINE database

The user would now be greeted by a registration page (see Fig. 2) where the user has to fill in details such as their name, email and affiliation. The form is automatically processed. Once the prospective user has filled in all required fields, an inactive account will be created for them, and they will automatically receive an email informing them of this. To activate the account, the prospective users need to sign the End User License Agreement (EULA, see section 4) and send the signed EULA to the email address provided. Information about this process is given both on the web page and in the emails that are send to the prospective users once they have completed the registration form. Notifications of account creations and account activation requests are sent to the database administrators. Normally they will not act on account creation notifications, but they will react to account activation requests. An account will be activated if the prospective user has correctly signed the EULA and meets the user criteria (i.e. the users are clearly academics).

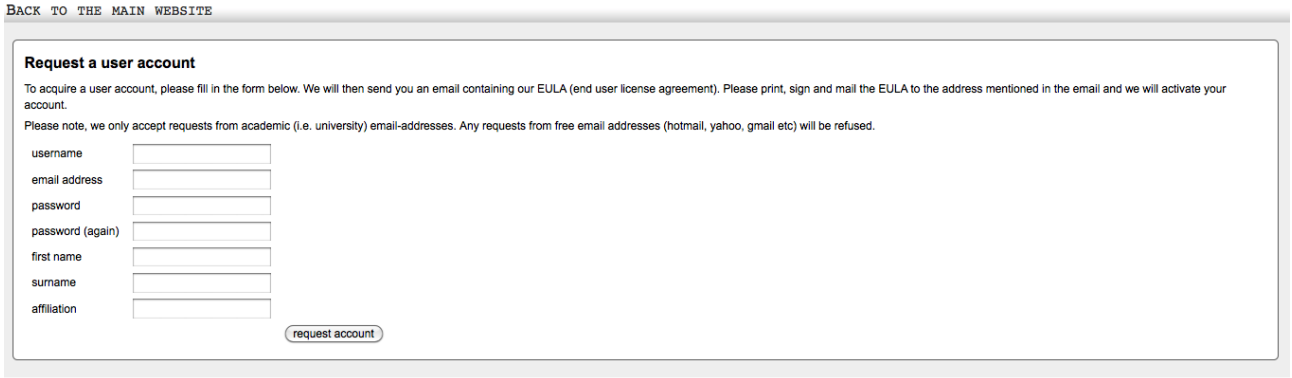

Figure 2: Registration form

Because of the EULA requirements, anyone who tries to register with an email who's domain name ends with 'hotmail.com', 'gmail.com', 'yahoo.com', etc. will be notified that users can only register using addresses issued by academic institutes.

3.2 Organisation of data

Within the database, the data is organised in units that we call a 'Session'. A session is a part of a recording, in which the User speaks with a single Character. There are also two special sessions per recording, to wit, the 'recording_start' and 'recording_end' sessions. These sessions include footage of the user preparing to do the experiment, or ending the experiment. Although these sessions do not show the desired User/Character interaction, they may still be useful for training modules that do not need interaction, such as the facial point detection module, or the User Presence module.

Each Session has a number of sensor data files associated with it. We call these database entries 'Tracks'. The Tracks of a Session consist primarily of the five camera recordings and the four microphone recordings (see Section 2.2). In addition, each Session has two lower-quality audio-visual Tracks, one showing the frontal colour recording of the User, and the other showing the frontal colour recording of the Operator. Both low-quality recordings have audio from the Operator and the User. To allow annotators to focus on only one person talking, we stored the User audio in the left audio channel, and the Operator audio in the right audio channel. Because most media players have a 'balance' slider, an annotator can choose who to listen to.

In our database, all annotation files (Annotations) are associated with a Track. It is possible that a single annotation belongs to multiple tracks: for instance, the affective state of the User should be associated with all Tracks that feature the User. Other Annotations can be associated with only a single Track.

Sessions, Tracks, and Annotations are displayed conveniently in a tree-like structure. One can click on the triangles in front of tree nodes to view all branches. Apart from the Tracks and Annotations, each Session also shows information of the people that are present in the associated recording. This information about the persons shown is anonymous: it is impossible to retrieve a name of the subject from the database. In fact, this information is not even contained in the database. The database organisation is illustrated in Fig. 3.

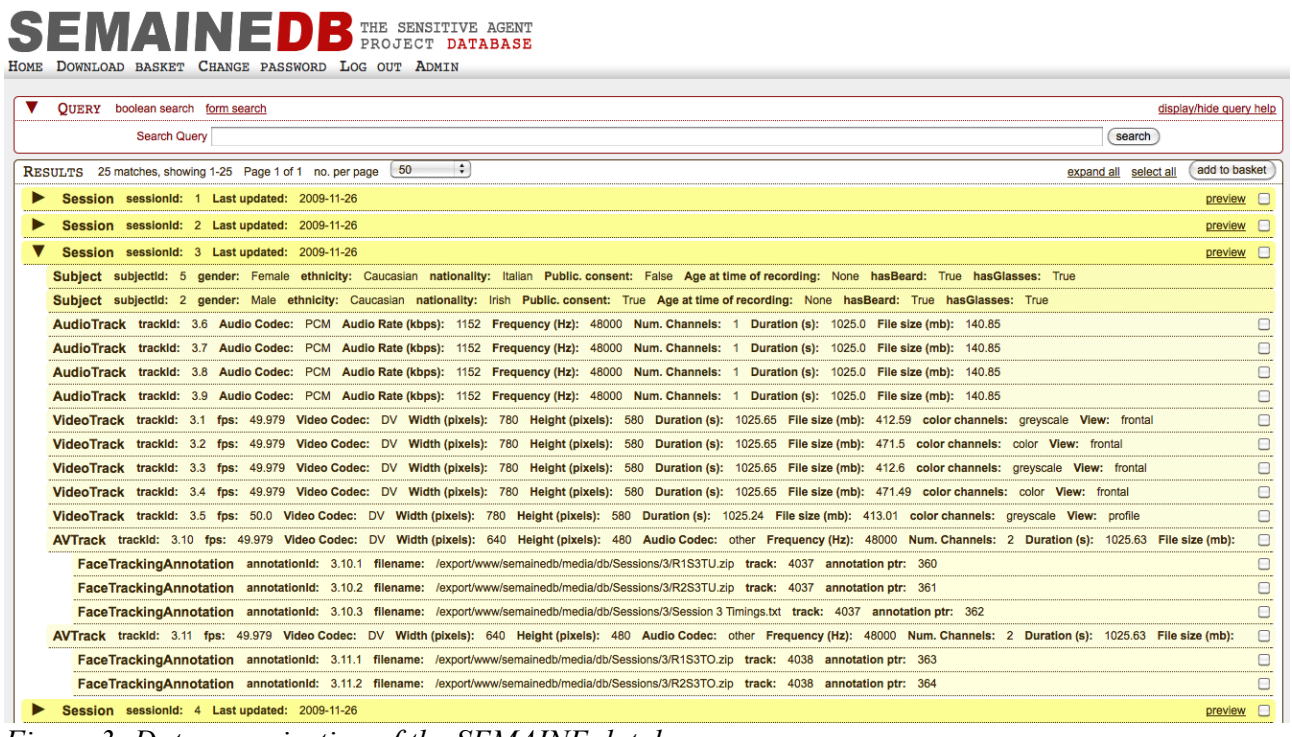

Figure 3: Data organisation of the SEMAINE database

3.3 Search functionality

To allow researchers to conveniently find the data they require, we have implemented extensive database search options. Searching the database can be done in two different ways: either by using regular expressions or by selecting elements to search for in a tree-structured form. The regular expression search is mainly intended for people who work with the database on a day to day basis and who know the search options by heart. Ordinary users are expected to use the form based search.

Search criteria can use characteristics of Sessions, Subjects, Tracks, and Annotations. It is possible to search by User gender, age, and nationality, by Session Character, by active AUs, and many many more. Figure 4 gives an example of the search form.

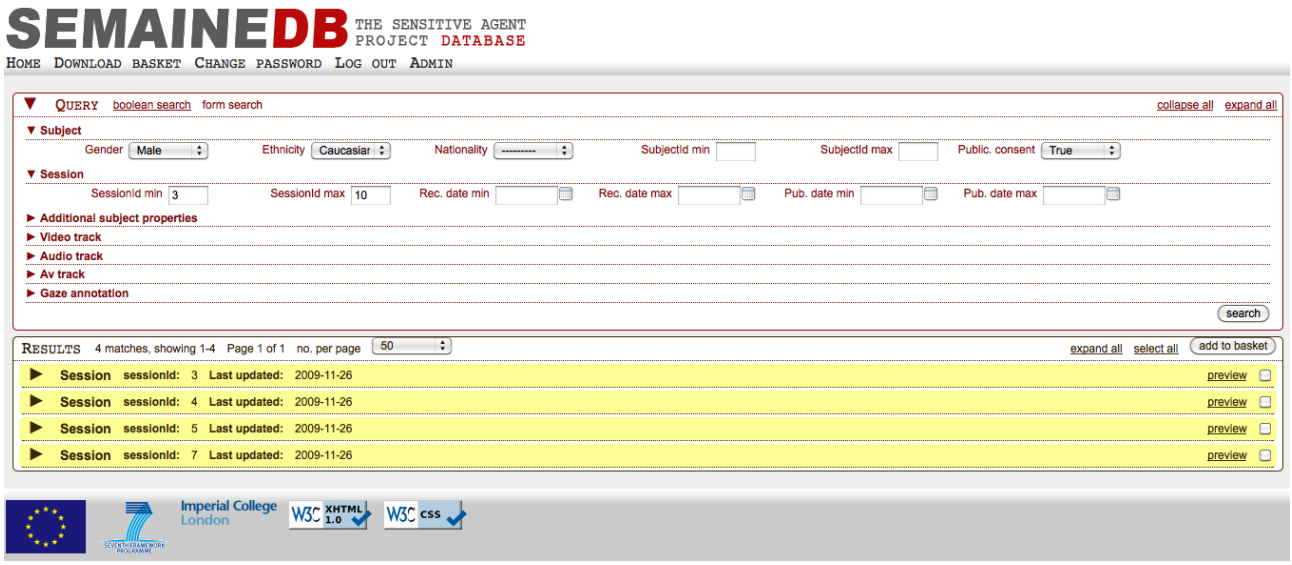

Figure 4: Form search: some options and the search results.

Once the user has executed a search and found elements of the data they like, they can check the tickboxes of these data elements and add them to their download basket. They can continue making searches and adding (parts of) the search results to their download basket until they are satisfied that they've got everything they need. The user can then proceed to the download section. Clicking 'download' will generate a single compressed zip file containing all the files in the download basket, which can then be downloaded conveniently.

4 License and availability

The EULA allows access to the database to anyone who wishes to conduct scientific, non-commercial research with it. It also states that the creators and providers of the database can not be held liable for damages caused by using the database. It states that anyone who wishes to use the database in their scientific publications must cite a specific paper that describes the database. It also states that users of the database cannot re-distribute the data.

The full EULA is provided as an Appendix to this document.

References

[Bales1951] Bales, R. F. (1951). Interaction process analysis: A method for the study of small groups. Cambridge, MA: Addison-Wesley Press.

[Baron-CohenEtAl2004] Baron-Cohen, S., Golan, O., Wheelwright, S. & Hill, J. J. (2004) Mind Reading: the interactive guide to emotions.London, UK: Jessica Kingsley Limited. (www.jkp.com/mindreading)

[Douglas-CowieEtAl2007] E. Douglas-Cowie, R. Cowie, I. Sneddon, C. Cox, O. Lowry, M. McRorie, J.C. Martin, L. Devillers, and A. Batliner. "The humaine database: addressing the collection and annotation of naturalistic & induced emotional data", 2nd International Conference on Affective Computing and Intelligent Interaction, LNCS, vol. 4738, pp. 488-500, 2007

[Ekman1999] Ekman, P. (1999). Basic emotions. In T. Dalgleish & T. Power (Eds.), The hand- book of cognition and emotion (pp. 45–60). Sussex, UK: John Wiley.

[FontaineEtAl2007] Fontaine, Scherer, Roesch & Ellsworth. "The World of Emotions Is Not Two-Dimensional". Psychological science vol. 18 (2) pp. 1050-1057, 2007

[LichtenauerEtAl2010] J.F. Lichtenauer, J. Shen, M.F. Valstar, M. Pantic, "Cost-Effective Solution to Synchronised Audio-Visual Data Capture using Multiple Sensors", Journal of Visual Communication and Image Representation, accepted for publication

Appendix I – End User License Agreement

EULA – End User License Agreement SEMAINE database (http://www.semaine-db.eu).

This user agreement applies to the Semaine database, first provided on the 1st of January 2010. It has been released by the Semaine consortium under specific conditions for sole scientific, non-commercial use by the Semaine consortium. The data was recorded for Semaine on the site of QUB (Queen's University Belfast). The database is stored and maintained by the HCI^2 group at ICL (Imperial College London School of Medicine and Technology). By signing this document the user, he or she who will make use of the database or the database interface, agrees to the following terms. In this document, the term 'database' denotes both the actual data and the interface to the data.

1. Commercial use
The user may not use the database for any commercial purposes. Commercial purposes include, but are not limited to:

- proving the efficiency of commercial systems,
- testing commercial systems,
- using screenshots of subjects from the database in advertisements,
- selling data from the database

2. Distribution
The user may not distribute the database in any way. Small portions (screenshots) may be distributed in publications as long as the publication complies with the terms stated in this EULA (article 4).

The user will forward all requests for copies of the database by third parties to the Semaine database administrators.

3. Access
The user may only use the database after this EULA has been signed and returned to the HCI^2 Group at Imperial College London. The user must return the signed EULA by email, in PDF format to the following address:

eula@semaine-db.eu

Multiple users may sign one EULA in order to grant access to a group of researchers. However, each user (researcher) must register with an individual username and password.

The user may not grant anyone access to the database by giving out their username and password.

4. Publications
Publications include not only papers, but also presentations for conferences or educational purposes.

The user may use screenshots of subjects in publications only if that particular subject has explicitly granted the permission to use his or her recordings in publications. The database specifies whether a user has given this consent to use their imagery in publications or not.

All documents and papers that report on research that uses the Semaine Database will acknowledge this as follows: "(Portions of) the research in this paper use the Semaine Database collected for the Semaine project (www.semaine-db.eu)".

The user will send an e-copy of all papers that reference the database to:

publication@semaine-db.eu

5. Research
The user may only use the database for scientific research, teaching, and/or classroom use. Any models derived using data from the dataset may only be used for scientific, non-commercial applications.

6. Changes
The Semaine consortium is allowed to change this EULA at any time; users will be informed about changes beforehand and given the choice to opt out of the new EULA. Opting out will render the previous EULA void.

7. Warranty
The database comes without any warranty, the HCI^2 Group at Imperial College London can not be held accountable for any damage (physical, financial or otherwise) caused by the use of the database. No legal claims of any kind can be derived from accepting and using the Dataset. The HCI^2 Group at Imperial College London will try to prevent any damage by keeping the database virus free.

I have read and understood the user agreement and will comply with it.

Signed_____

Please print name_____

Institution_____

Date_____