

humaine

D5i: Final report on WP5

Workpackage 5 Deliverable



Date: 30th January 2008

IST project contract no.	507422
Project title	HUMAINE Human-Machine Interaction Network on Emotions
Contractual date of delivery	<i>January 30, 2008</i>
Actual date of delivery	<i>February 14, 2008</i>
Deliverable number	D5i
Deliverable title	Final report on WP5
Type	Report
Number of pages	12
WP contributing to the deliverable	WP 5
Task leader	QUB
Author(s)	E.Douglas-Cowie
EC Project Officer	Philippe Gelin

Address of lead author: Ellen Douglas-Cowie
 Dean of Arts, Humanities & Social Sciences
 Queen's University
 BELFAST BT7 1NN
 Northern Ireland

TABLE OF CONTENTS

1	THE PLACE OF THIS REPORT WITHIN HUMAINE.....	4
2	ACHIEVEMENTS	5
2.1.1	Basic records	5
2.1.2	Labelling.....	6
2.1.3	The HUMAINE database	8
2.1.4	Conclusion and additional sources.....	11
3	WP5 DELIVERABLES (ALL OF THESE ARE AVAILABLE ON THE PORTAL)	
	12	

1 The place of this report within HUMAINE

The final workpackage reports in HUMAINE have been designed as brief statements that complement other sources. On one hand, substantive content – particularly technical content – has been reported in deliverables throughout the project. On the other hand, several potentially overlapping sources are part of the final reporting process. In particular, the activity report for the final period sets out the goals achieved during the final period; and the final report for the project as a whole includes a relatively non-technical account of what the workpackage has achieved.

This core of this report is shared with the section on WP5 in the final report for the project as a whole. That consists of a short, non-technical summary of what the workpackage has achieved. However, there are details that do not belong in a report intended for the general reader (such as information about the participating institutions, meetings organised, etc). They are covered here.

Workpackage members invested a substantial amount of effort in the deliverables, and they form a continuing resource available to the community on the HUMAINE portal. They are listed for completeness. In some senses it would make sense to do the same for publications, but they are too difficult to allocate to individual workpackages. A full list is given in the final report for the project as a whole.

2 Achievements

Achievements in WP5 can be divided into two broad areas, collection of basic records and labelling. They have been integrated in a database which is available on the portal.

2.1.1 Basic records

The database team collected a highly varied set of records. Almost all were multimodal, and either naturalistic or induced in a way that produces relatively natural behaviour.

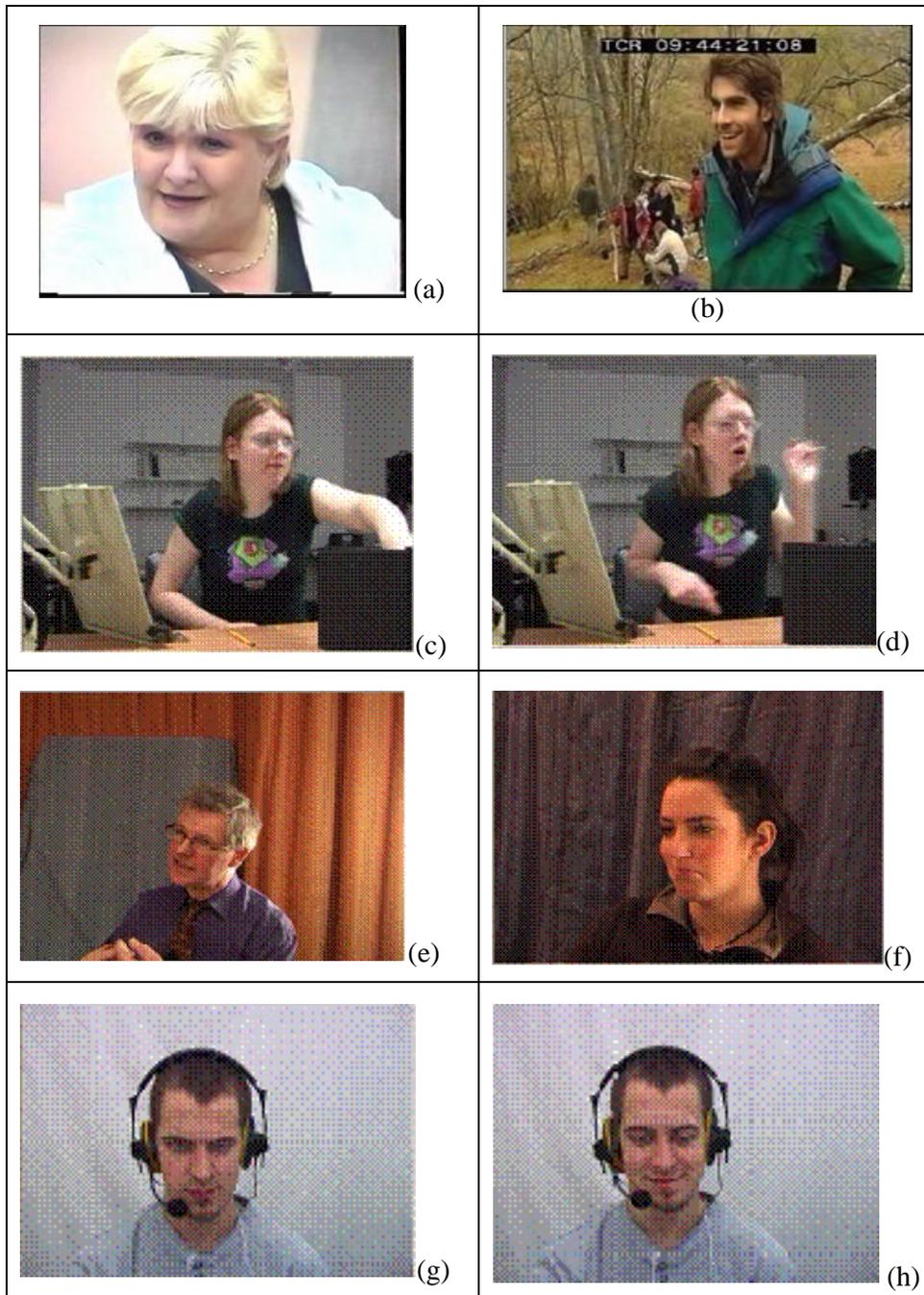


Figure 1: Illustrative data from datasets used by WP5: (a) interview between friends from *Belfast Naturalistic Database*; (b) outdoor activity from *Castaway Reality Television Dataset*; (c) & (d) laboratory task eliciting startle from the *Spaghetti Dataset*; (e) speaker attempts to persuade listener (f) to adopt a ‘green lifestyle in the *Green Persuasive Dataset*; (g) & (h) participant in simulated driving task with & without cognitive load from *DRIVAWORK (Driving under Varying Workload) corpus*(in German).

The variety of the material is illustrated in Figure 1. Several other datasets were also collected or used in association with the project. Recordings were made using the *Sensitive Artificial Listener* paradigm (the speaker interacts with an experimenter who simulates an ECA) in English, Greek and Hebrew. The *EmoTV Database* (extracts from news broadcasts) and *EmoTABOO* (a game that elicits complex gestures) were in French. All of the above provided audiovisual data. *DRIVAWORK* and the *Belfast Driving Simulator Dataset* also provided physiological measures and instrumentation of driving behaviour.

With a few exceptions (involving copyright problems), the records mentioned above will be made available to the research community in full. The corpus will give the community access to a much more varied, naturalistic and multimodal body of material than it has previously had access to.

2.1.2 Labelling

The team developed labelling techniques that are suited to capturing the phenomena that appear in recordings of relatively natural emotional interactions. Work has been done on three levels of description: perceived emotional content; the signs that convey the emotion; and the contextual factors relevant to understanding the signs.

The team developed and explored a wide range of ways to describe perceived emotional content. The default method is marking on a timeline when particular everyday emotion categories appear to apply. The team studied the reliability of that method, and explored various related techniques: prespecifying a list of terms to be used, drawing up ‘cover classes’, and combining results in ‘soft vectors’.

FEELtrace (developed by HUMAINE members) is a radical alternative. In the original version, users ‘trace’ perceived emotional state in terms of the two longest established dimensions, valence and activation, by moving a cursor within a space whose axes are valence (positive to negative) and activation (highly energised to deeply inert). The idea of using a continuous trace to report perceived flow of emotion is quite powerful, but tying it to valence and activation is limiting. Hence the team built on the trace concept by developing a wide range of one-dimensional traces, each dealing with a dimension that either theory or intuition indicate might be significant in the perception of emotion. Tests on the reliability of a core set were carried out.

Table 1 below summarises traces that appeared reliable enough to be incorporated into the HUMAINE database procedure (described in the next section). The technique can very easily be extended, and some other interesting variants have been used but not checked in depth. CatTrace records the broad kind of emotional state a person is in (emotion in the strong sense, truly unemotional, intermediate). PersuadeTrace was applied to the *Green Persuasive Dataset* (see above): persuadees used it to record how persuasive they had found the speaker from moment to moment.

Traces were among several types of technique explored in an attempt to develop labellings based on appraisal theory. In general, labelling appraisal proved difficult. The basic reason is

that appraisals involve an object (e.g. ‘goal-obstructiveness’ is an attribute of a particular thing or event in the person’s surroundings or mindscape, not the whole environment). Identifying the object as well as the orientation to it is a different level of task, and the best ways to do it are not yet clear.

Table 1: Trace-type descriptors developed by HUMAINE teams and used in the HUMAINE database

Trace label	Which attribute of X (the person whose state is being traced) is rated
IntensTrace	The perceived intensity of the emotion that X is experiencing from moment to moment.
ActTrace	The perceived extent to which X is trying to give an impression of emotion that he/she does not actually feel.
MaskTrace	The perceived extent to which X is trying to avoid showing emotions that he/she does actually feel.
ActivTrace	X’s perceived level of activation or arousal.
ValenceTrace	X’s perceived positive or negative feeling about the events or people at the focus of his/her emotional state.
PowerTrace	X’s perceived sense of control over the events or people at the focus of his / her emotional state.
ExpectTrace	X’s perceived sense of having anticipated or been taken unawares by the events at the focus of their emotional state.
WordTrace	The extent to which X’s emotional state appears to involve the emotion specified by a given word (anger, happiness, etc).

A range of techniques for describing signs of emotion was also developed, often in conjunction with other workpackages. Gesture description techniques were developed in conjunction with WP6, working on synthesis, so that the descriptions in databases would relate as directly as possible to the task of synthesis (to the extent that a joint paper was presented describing copy synthesis from an example in the *EmoTV Database*). Descriptions of both some aspects of gesture and facial features were agreed with WP4, so that descriptions could be extracted automatically. In addition, a system for describing emotion-related features of speech was developed.

Systematic description of context is not a new goal, but its importance has not been systematically recognised. An example used in HUMAINE discussions was that newly married couples are likely to signal their feelings in very different ways before and after their in-laws come into the room. The point is that an automatic system cannot be expected to understand the signals unless it is attuned to the way they depend on context. The context labels that were developed separated three broad types of issue. ‘Constraint’ covers both physical limitations (e.g. seated posture limits movement) and social pressure (e.g. appearing

before a military tribunal – which participants in some databases actually were). ‘Goal’ covers various communicative functions – to inform, to create rapport, etc. ‘Setting covers the kind of structure within which the communication takes place – shared activity, dialogue, group discussion, etc.

2.1.3 The HUMAINE database

The HUMAINE database is a collection of all the records described under the ‘Records’ section and a labelled subset.

HUMAINE was not resourced to label all of the records that members assembled. The strategy chosen was to label a relatively small subset with two aims in mind:

- to reflect the variety of material that is potentially relevant to emotion-oriented computing;
- to demonstrate the variety of techniques that may be needed to label emotionally coloured material in a satisfying way.

48 clips (between 3secs and 2 minutes in length) were selected for labelling from the primary recordings. Table 2 provides a summary of the numbers of clips selected from the range of data types to make up the labelled subset.

Table 2: Selection of clips for labelled subset

Raw Data Records	Data type	Number of clips selected
<i>Belfast Naturalistic Database</i>	naturalistic	10
<i>Castaway Reality Television Dataset</i>	naturalistic	10
<i>Sensitive Artificial Listener (Belfast recordings: English)</i>	induced	12
<i>Sensitive Artificial Listener (Tel Aviv recordings: Hebrew)</i>	induced	1
<i>Activity Data/Spaghetti Data</i>	induced	7
<i>Green Persuasive Dataset</i>	induced	4
<i>EmoTABOO</i>	induced	2
<i>DRIVAWORK (Driving under Varying Workload) corpus: German</i>	induced	1
<i>GEMEP Corpus</i>	acted	1

The 48 clips were deliberately selected to cover material showing emotion in action and interaction; in different contexts (static, dynamic, indoor, outdoor, monologue and dialogue); spanning a broad emotional space (positive and negative, active and passive) and all the major types of combination of emotion (consistent emotion, co-existent emotion, emotional transition over time); with a range of intensities; showing cues from gesture, face, voice, movement, action, and words and representing different genders and cultures.

The labelled episodes are mounted on the ANVIL platform. Labelling is at two levels.

At the first level, global labels are applied to an emotion episode or clip as a whole. This provides a natural way of describing factors that do not vary rapidly (the person concerned,

the context) are described here. It also provides an index that can be used to identify clips that a particular user might want to consider (for instance, how anger is expressed in relatively formal interactions).

Labelling at the second level is time-aligned. Emotional content is labelled using ‘trace’ type programs. Different signs use different formats. Figure 2 illustrates how key elements appear – speech (raw waveform and transcription); emotion-related traces; gestures and facial expression.

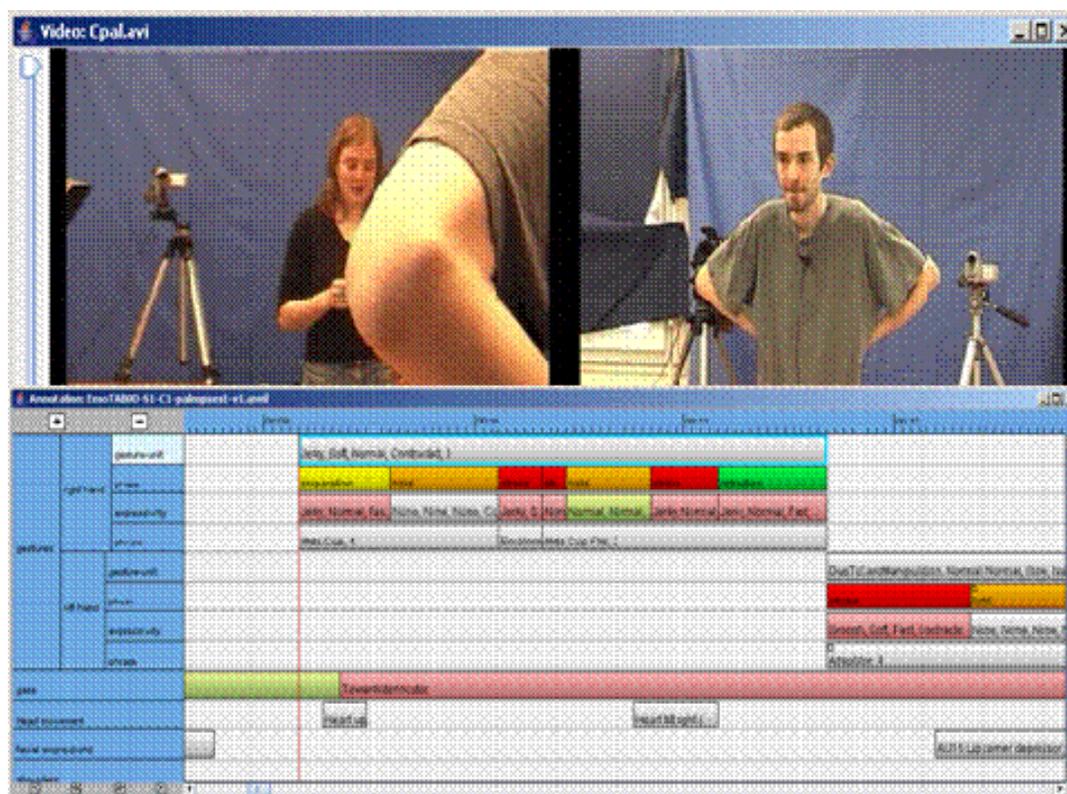
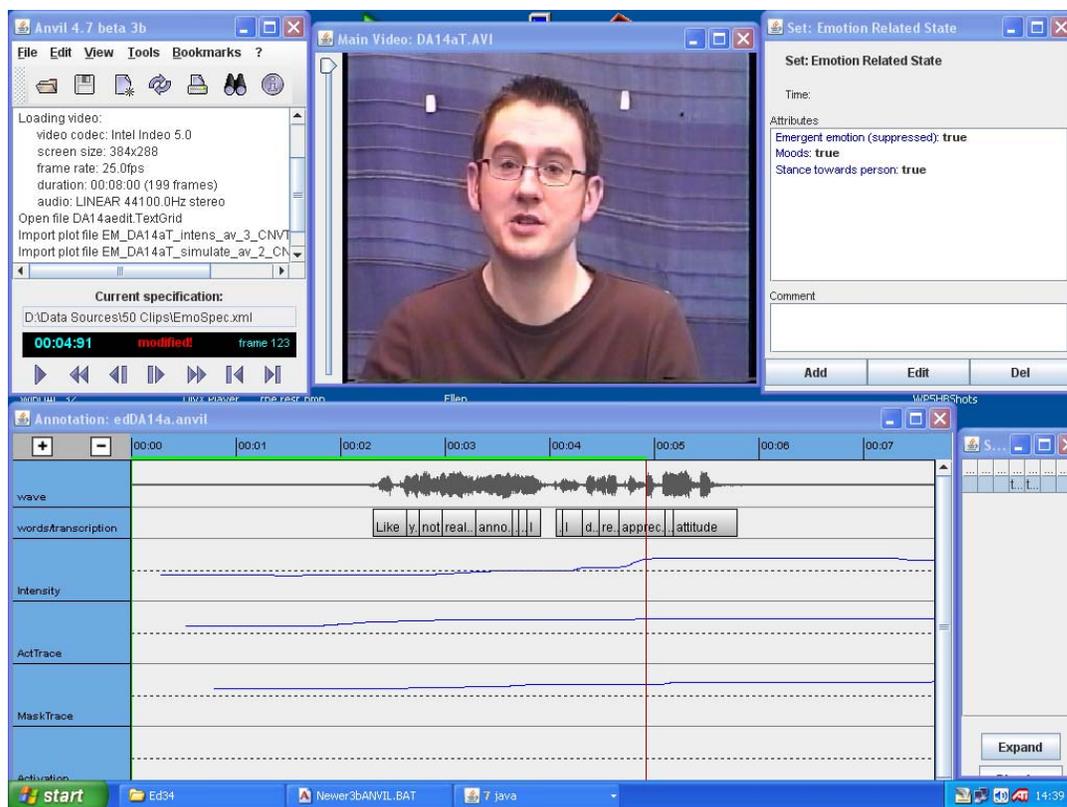


Figure 2 : Clips from the HUMAINE database showing ANVIL annotation for emotion (top panel) and gesture (bottom panel)

2.1.4 Conclusion and additional sources

When plans for HUMAINE began, there was a growing recognition in the field that a major obstacle to developing accurate models of human emotion was the absence of rich, realistic databases. Researchers were often dependent on small, ad hoc datasets, or on their own intuitions, or on data that focused on archetypal, isolated examples of emotion, usually acted. Data was often unimodal.

WP5 has made a substantial contribution to changing that scene. It has:

- provided a library of recordings that can underpin meaningful research into emotion as it appears in action and interaction in everyday life including
 - recordings that cover signs in visual, audio, and physiological modalities
 - recordings that provide data across a range of cultures
- developed and disseminated techniques for eliciting emotion as it appears in everyday life including
 - techniques for eliciting a wide range of emotional behaviour
 - techniques for eliciting human-human and human-machine emotional interaction
 - techniques for eliciting emotion in action and interaction
 - techniques for eliciting a wide range of signs of emotion
- shaped and disseminated descriptive schemes for labelling including
 - labels for the perceived emotion at both a global level and in continuous ‘trace’ form over time
 - labels to describe emotional context
 - labels for signs of emotion
- provided a fully labelled subset (mounted on the ANVIL platform) consisting of
 - data from 6 raters for the labelled subset
 - manuals describing the different labelling schemes and on how to use them
 - the labelling tools (the programs in the case of Trace labels)

Finally, WP5 members have made the ‘know how’ developed widely available to researchers in the field by uploading the HUMAINE database on the portal. This contains, not just the means to access the records themselves, but also the labelling tools and programs and papers describing the elicitation techniques. In the course of HUMAINE WP5 members have also led the way in emotional database research by organising three successful workshops/special sessions at major international conferences (the Interspeech 2005 workshop on Emotional Speech which was organised by members from QUB and CNRS-LIMSI and had a strong emphasis on databases, the workshop specifically on Corpora for Research on Emotion and Affect in LREC 2006, and organised by members from QUB, CNRS-LIMSI and FAU-Erlangen and the upcoming workshop on this theme at LREC 2008). Over and above this, members of WP5 have given and published key papers individually and on a team basis. Over the period of HUMAINE they have responded to many requests from across the world for emotional databases and for labelling tools.

3 WP5 Deliverables (all of these are available on the portal)

Del. no	Deliverable name
D5a	Workshop proceedings: Databases
D5b	Collection of existing databases
D5c	Preliminary Plans for Exemplars - Databases
D5d	potential exemplars: databases
D5e	Proposal for exemplar and work towards it: Data and Databases
D5f	Pilot collection of labelled naturalistic and induced material available via portal
D5g	Mid term report on database exemplar progress
D5h	Database Exemplar Mid Term Report
D5i	Final HUMAINE Database