The HUMAINE database

Ellen Douglas-Cowie, Cate Cox, Jean-Claude Martin, Laurence Devillers, Roddy Cowie, Ian Sneddon, Margaret McRorie, Catherine Pelachaud, Christopher Peters, Orla Lowry, Anton Batliner, Florian Hönig

Angaben zur Veröffentlichung / Publication details:

Douglas-Cowie, Ellen, Cate Cox, Jean-Claude Martin, Laurence Devillers, Roddy Cowie, Ian Sneddon, Margaret McRorie, et al. 2011. "The HUMAINE database." In *Emotion oriented systems: the humaine handbook*, edited by Roddy Cowie, Catherine Pelachaud, and Paolo Petta, 243–84. Heidelberg [u.a.]: Springer. https://doi.org/10.1007/978-3-642-15184-2_14.

The HUMAINE Database

Ellen Douglas-Cowie, Cate Cox, Jean-Claude Martin, Laurence Devillers, Roddy Cowie, Ian Sneddon, Margaret McRorie, Catherine Pelachaud, Christopher Peters, Orla Lowry, Anton Batliner, and Florian Hönig

Abstract The HUMAINE Database is grounded in HUMAINE's core emphasis on considering emotion in a broad sense – 'pervasive emotion' – and engaging with the way it colours action and interaction. The aim of the database is to provide a resource to which the community can go to see and hear the forms that emotion takes in everyday action and interaction, and to look at the tools that might be relevant to describing it. Earlier chapters in this handbook describe the techniques and models underpinning the collection and labelling of such data. This chapter focuses on conveying the range of forms that emotion takes in the database, the ways that they can be labelled and the issues that the data raises. The HUMAINE Database provides naturalistic clips which record that kind of material, in multiple modalities, and labelling techniques that are suited to describing it. It was clear when the HUMAINE project began that work on databases should form part of it. However there were very different directions that the work might have taken. They were encapsulated early on in the contrast between 'supportive' and 'provocative' approaches, introduced in an earlier chapter in this handbook. The supportive option was to assemble a body of data whose size and structure allowed it to be used directly to build systems for recognition and/or synthesis. The provocative option was to assemble a body of data that encapsulated the challenges that the field faces.

The eventual choice leant heavily towards the provocative. The supportive was not wholly ignored. Partners in the HUMAINE Database work package have developed the resources used in several significant projects on recognition of emotion, both unimodal (CEICES) and multimodal (Cowie et al.). However, the main systematic effort was directed towards establishing a corpus that summed up the challenges facing the community and drew together key resources that are potentially relevant to meeting them. The choice is grounded in HUMAINE's core emphasis on

E. Douglas-Cowie (✉)
Department of Psychology, Queen's University Belfast, Belfast, Northern Ireland, UK
e-mail: E.Douglas-Cowie@qub.ac.uk

considering emotion in a broad sense – 'pervasive emotion' – and engaging with the way it colours action and interaction. Until now, there has been no source to which the community could go to see and hear the forms that emotion takes in everyday action and interaction, and to look at the tools that might be relevant to describing it. The core aim of the HUMAINE Database was to provide that kind of source. The description in this chapter reflects that orientation. It does not set out to give technical specifications of the database contents. Instead it sets out to convey the range of forms that emotion takes in the database and the ways that the descriptive resources address them. The database is not simply about creating impressions. It is also designed to let key technical questions be addressed. That aspect is taken up at the end of the chapter.

1 Overview and Structure

This section sets out some of the specific concerns underlying the database.

The emphasis is on multimodal data. Most of the material is audiovisual. The visual channel usually includes face, but some sources deliberately contain gesture and some include body posture or mode of action. Other parts include physiological data and choice of response to challenges.

The material was collected to show a wide emotional range (negative to positive, active to inert, deeply engaged to playful). The emotion is embedded in a range of actions and interactions, and a variety of contexts.

The material makes varying levels of allowance for the limits of contemporary signal processing. At one extreme is data from TV recordings, shot outdoors with 'difficult' camera angles and noisy audio; at the other is data derived from laboratory scenarios devised to minimise signal processing challenges.

The labels that are used are described in the preceding chapter of this part. They include labels describing emotional content, based on the psychological literature; labels for emotional signs in the relevant modalities (face, speech and gesture context labels have been developed) and context labels. The labels span a range of resolutions in time (whole passage to moment by moment).

There are close links between the recordings and the labels. On the one hand, labels were chosen to deal with the phenomena that were observed in the recordings that laid the groundwork for the database. On the other hand, material was generated and selected to reflect the range of possibilities that the labels indicated ought to be represented.

The database consists of two parts: (i) primary records and (ii) a structured labelled subset. The primary records consist of recordings, almost all audiovisual, in diverse emotion-rich scenarios. The structured labelled subset is a balanced and labelled set of emotional episodes (referred to as 'clips') selected from the primary records to represent a range of emotions and to demonstrate the application of a wide range of labels covering emotional content, context and signs. Most of the primary

records and the whole of the labelled subset are available to the research community under 'Conditions of Use Agreement' (see Appendix 1 for details).

The primary records consist of many hours of recordings and contain emotional episodes which have not been identified and selected for labelling. They are thus a resource which can be mined by other researchers. They are also a useful resource for observing how often emotionality of some level occurs across time. Some of the data are 'naturalistic' (in the sense that it has been collected in situations not under the researcher's control, e.g. from film shot for television, etc.). Some are 'laboratory'-induced data (in the sense that emotionality has been induced in a controlled environment according to a purpose-specific method). The laboratory-induced data is a rich resource, not just for the data itself but also for the range of methods used to carry out the induction of emotion (developed specially for the HUMAINE project – for fuller details, see chapter "Issues in Data Collection").

One of the principles behind the HUMAINE database was that the data should be available to the community in general. With the particular nature of the data, this means that there have been important issues of ethical clearance and consent to be addressed. The ethical principles underpinning ethical clearance and consent are dealt with more fully in Part IV 'Ethics and Good Practice'. Access to the database is via the HUMAINE Association portal (www.emotion-research.net).

2 The Total Data Set

The HUMAINE Database work package recorded or acquired a large body of material showing emotion as it appears and sounds in action and interaction. This section summarises the main kinds of material that have been collected as a result.

2.1 Summary of Primary Records

Table 1 provides a summary of the data types that make up the primary records. The recordings are usually either naturalistic or induced, although an emotional episode is selected from one professionally acted data set (GEMEP, see Baenziger and Scherer, 2007) for labelling (as a comparison). The GEMEP Data Set as a whole is not available as part of the HUMAINE database. The induction techniques that have been developed to induce much of the material are described in chapter "Issues in Data Collection" in this part. Appendix 2 describes in more detail the exact nature of the primary records – the material, technical information about the material including length, numbers of subjects, recording scenario and the conditions under which it is available. Appendix 3 describes in detail each episode or clip selected for labelling (to form part of the labelled subset) including a summary of the content of the emotional episode and descriptors of the emotion, context and modalities.

Table 1 Data types used in the HUMAINE Database

Primary records	Data type
Belfast Naturalistic Database	Naturalistic/induced: television data and interviews between friends
EmoTV Database (in French)	Naturalistic
Castaway Reality Television Data Set	Naturalistic
Sensitive Artificial Listener (Belfast recordings in English)	Induced
Sensitive Artificial Listener (Tel Aviv recordings in Hebrew)	Induced
Activity Data/Spaghetti Data	Induced
Green Persuasive Data Set	Induced
EmoTABOO	Induced
DRIVAWORK (driving under varying workload) corpus in German	Induced
GEMEP Corpus (Geneva Multimodal Emotion Portrayal)	Acted
Belfast Driving Simulator Data Set	Induced

2.2 Illustration of Data Types

This section illustrates the nature of the data, pulling out its typical characteristics, theoretical interest and relevance and its strengths and weaknesses. It starts with the naturalistic data.

Figure 1 shows a typical frame taken from the Belfast Naturalistic Database (Douglas-Cowie et al., 2003). The subject is talking to an old friend about how she feels about her future son-in-law and expressing her positive feelings for him. The frame is typical of the data in that the emotion expressed is strong but not full blown. In terms of the emotion-related states described in the first chapter in this part, the emotion expressed in this particular example reflects a long-lasting feeling ('attitude') towards someone or something in the Belfast Naturalistic Database. Much of the data either expresses attitudes or is 'established' emotion (long-standing states that can be 'triggered' in a way that produces surges of overt emotion). In this example the subject is sedentary and the camera is fixed on the face, head and shoulders.

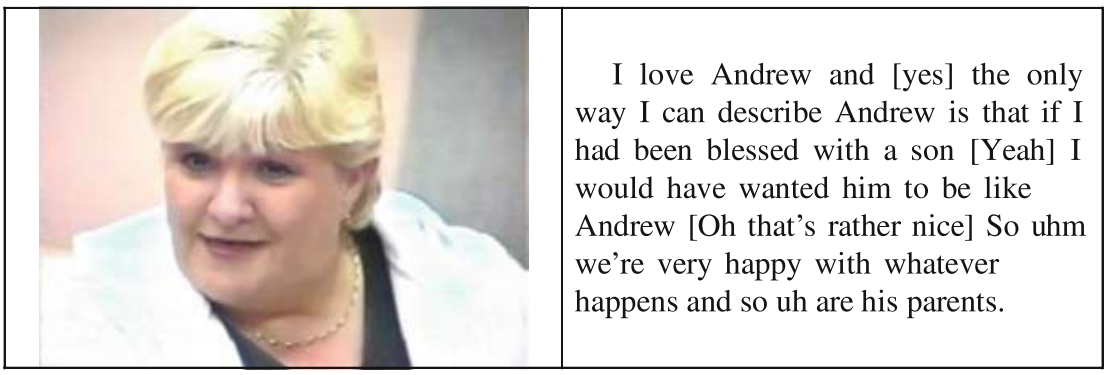

Fig. 1 Frame from Clip56b, Belfast Naturalistic Database

The visual quality is not perfect. The induction technique captures a lot of speech in a dialogue situation and because the old friend is also the researcher she knows not to interrupt too much, thereby giving long stretches of uninterrupted speech for analysis. The text is given beside the picture with the interviewer's comments in square brackets.

Figure 2 shows two frames, both from the Castaway Reality Television Data Set (Douglas-Cowie et al., 2007). The data set contains some intense emotion, and these frames illustrate that. The one on the left is of a subject in a positive state (after successful completion of a task) and the one on the right is of a subject in a negative state (after a task in which he thinks he has done badly). The material is important because it shows emotion in action as participants engage in a range of challenging activities both singly and in groups. It often shows shifting and complex emotions as a subject moves through a challenging activity and comes out at the end successfully or unsuccessfully. Because there is a lot of activity, there is a lot of movement by the participants and this gives rise to data that is not face-on or close-up, and so from an affective computing point of view, the data is quite challenging. The shots in Fig. 2 illustrate some of the more static material in the data. Most of the speech occurs in the one-to-one interactions with the team leader, but outdoor noises from the rest of the group tend to get in the way of good recordings. Nevertheless, pitch traces have been reliably extracted for the clips from this data that form part of the labelled subset.

Figure 3 shows two frames from the SAL (Sensitive Artificial Listener) induction technique (see "Issues in Data Collection", Part III). SAL involves an artificial character with different emotional personalities (Poppy who is sad, Spike who is angry, Obadiah who is gloomy, Prudence who is sensible) engaging a subject in conversation. Each personality uses stock responses and phrases to pull the subject towards his/her mood. In Fig. 3 the subject is shown talking to the gloomy personality of the artificial listener (on the left) and then the happy personality of the artificial listener (on the right). The frames show reasonably natural data which demonstrates mild to moderate levels of emotion. The subject is sedentary and the camera is focused on the face and shoulders. There is no gesture, but the technique generates a lot of

Fig. 2 Frames from the Castaway Reality Television Data Set (from Clip5_2, *left*, and Clip 6_2, *right*)

Fig. 3 Subject in conversation with two different personalities of the Sensitive Artificial Listener (from Clip REllA2, *left*, and Clip REllB2, *right*)

Table 2 Text accompanying SAL clips REllA2 and REllB2 (SAL personality responses in square brackets)

SAL clip	Text
REllA2	Tch, well, erm you used to be just gloomy but you seem to be getting logical now, that's a bit worrying [I don't suppose it really matters though] No it doesn't really. I mean, you don't have anything practical to offer at all, no help whatsoever. You just want to sink down into a gloomy state, but I'm not going to do that Obadiah. You go away, I'm, I'm going to talk to somebody like Prudence, she's pragmatic, she'll tell me what to do and then I'll feel a lot better
REllB2	Er, or, another thing that I really like and I haven't really managed to do it this year and I really really must do it but I'm too tired to do it but it's going out on the bike and going up the towpath on a summer evening. I'd love to do that at the minute and just sort of cycle up and all the hawthorne would be out and I could smell the hawthorne and you know it would be nice and kinda sunny and lovely air blowing across you. You know that nice feeling on a bike when the air kind of blows all around you. It's lovely

speech (see Table 2. The visual and auditory quality is good and SAL data has successfully been used to train an emotion recognition system (Ioannou et al., 2005).

The Belfast Activity/Spaghetti Data (see the chapter "Issues in Data Collection") is represented in the next set of figures (Figs. 4 and 5). Two techniques were used to produce the emotion in action seen in these two figures.

In the first (Activity Data), volunteers were recorded engaging in outdoor activities (e.g. mountain bike racing) in an effort to produce examples of full-blown emotion in action for which we would have consent and ethical clearance. This produced 'provocative' data, very dynamic, with subjects moving around. It also had a noisy sound track with affect bursts, but little speech. The data is demonstrated in Fig. 4, which shows sequenced frames from the subject watching one of the volunteers fall off a mountain bike in a 3-s episode. It demonstrates full-blown emotion and quite complex emotional shift. Some interesting work has been done on the data

Fig. 4 Sequenced frames from Belfast Activity Data at 0.56 s (*left*), 2.16 s (*middle*) and 3.00 s (*right*)

showing the speed of transition in facial movement in this data as opposed to acted emotional data (Sneddon and McRorie, 2006).

In the second technique, a more controlled environment was used where certain kinds of 'ground truth' could be established. It is called the Spaghetti method, because participants are asked to feel in boxes in which there were unpleasant objects (including spaghetti) and buzzers that went off as they felt around. They recorded what they felt emotionally during the activity. Figure 5 shows a typical data. The first frame shows the subject in the build up to the climax where a buzzer sounds when the subject locates the object in the box. The second shows the subject at the moment when the buzzer sounds. The data is of good quality both auditorily and visually, although there is very little actual speech – the sound track consists mainly of exclamations. In the clip from which the frames below are taken, the only words uttered are 'Oh Jesus' at the moment when the buzzer goes off.

Figures 6 and 7 represent a more recent move by the HUMAINE team towards experimentation with induced data that shows interaction between two subjects. This is particularly relevant to emotional synthesis. Figure 6 shows data from the Green Persuasive Data Set (see this chapter and the chapter "Issues in Data Collection") where complex emotions are linked to varied cognitive states and interpersonal signals. In the Green Persuasive Data Set, one person tries to persuade another on a topic with multiple emotional overtones (adopting a 'green' lifestyle). Figure 7 shows data from the EmoTABOO Data Set (Zara et al., 2007) where the emphasis is on generating gesture and where subjects interact mainly through gesture and body movement to explain and understand a taboo or an unusual word

Fig. 5 Belfast Spaghetti Data (building up to climax, *left*, and in response to buzzer, *right*)

Fig. 6 Persuader and persuadee's response in Green Persuasive Data Set (clips Ex2A and PT2a)

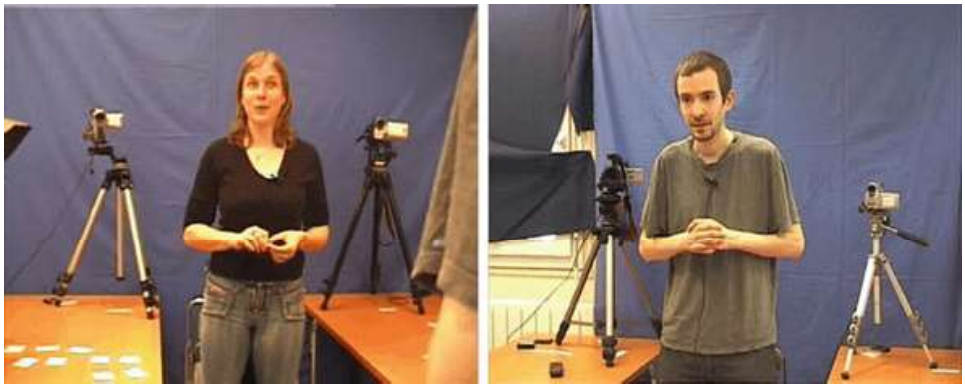

Fig. 7 Co-occurring frames from EmoTABOO, explainer on *left* and receiver on *right*

known only to one of the subjects. The scenario produces a lot of amusement and embarrassment.

Figure 6 shows the persuader on the left and the person he is trying to persuade on the right. The frame on the right is taken from immediately after the persuader's attempt to persuade the subject that cars are not needed in an environment-friendly 'green' world. She clearly disagrees. The data produces a lot of persuasive speech on the part of the persuader, less speech on the part of the persuadee but interesting facial responses. The text is given in Table 3. The quality both auditorily and visually is good.

Figure 7 shows an interaction between the subject who is trying to explain a 'secret' word/concept (on the left) to the person on the right who does not know the secret word. The two frames are from exactly the same moment. The data is very rich in gesture and is of good quality. There is some speech but the emphasis is on gesture.

The final picture in this section comes from the DRIVAWORK corpus (Honig, 2007). This uses a simulated driving task and subjects are recorded relaxing, driving normally or driving under an additional task (mental arithmetic). The technique aims to elicit physiological data in the different states. Figure 8 shows a subject relaxed (left), driving (middle) and under task (right). The task is context specific. The speech takes the form of answers to mental arithmetic problems. Speech and face are very clear.

Table 3 Exchanges between persuader and persuadee in clips Ex2A and PT2a

Persuader	Persuadee
Ehm, so I, I, I would tend to think, you know using public transport, sensibly, is.. is probably better than going to extremes	[Yea]
Extremes don't convince people	[Yea (pause) Er I do think anything like the new car is um really, like, electricity run, or um like um like]
Hydrogen?	[Yea Hydrogen]
Well, uh, ok, there, th.., there are three groups	[Mm Hm]
There's electric, uh, alcohol and hydrogen. Um, now, the electrical ones are basically a con because the electricity is being generated anyway, by and large being generated by coal, gas or other power stations	[Yea]
So it's, they're actually not very efficient, I mean it's uh, it's it's like one of these things where you know a child sort of, uh pretends they haven't eaten the sweets	[Laughs]
Just not, it's not really . . . ah. The, ehm, the pollutants they get rid of are the ones that you see. That's the the uh, the ah, the smoggy stuff	[Yea]
Ah, and sure that's not nice either but that's not the stuff that's killing the planet	[No]
So..	

The Belfast Driving Simulator Data uses specially developed induction techniques to record subjects driving in a range of emotional states. The procedure consists of inducing subjects into a range of emotional states and then getting them to drive a variety of 'routes' designed to expose possible effects of emotion. Induction involves novel techniques designed to induce emotions robust enough to

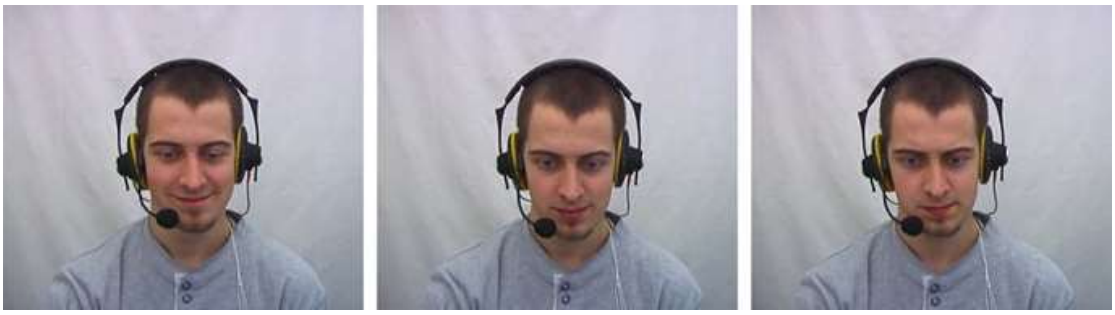

Fig. 8 Subject relaxed (*left*), driving normally (*middle*) and under task (*right*) taken from DRIVAWORK corpus

last through driving sessions lasting tens of minutes. Standard techniques are used to establish a basic mood, which is reinforced by discussions of topics that the participants have preidentified as emotive for them. The primary data is a record of the actions taken in the course of a driving session, coupled with physiological measures (ECG, GSR, skin temperature, breathing). It is supplemented by periodic self-ratings of emotional state. The data has not been video recorded. It is currently the topic of a Ph.D. and cannot be released until after completion of the Ph.D., but pilot work from the Ph.D. can be found at http://emotion-research.net/ws/wp5/edelle.ppt

3 The Labelled Subset

3.1 Aims of the Labelled Subset

The labelled subset is a balanced and labelled collection of extracts selected from the primary records. Each extract is selected to contain a relatively self-contained emotional episode and is referred to as a 'clips'. The subset was chosen to represent the range and form of emotional life that people working in the field should be aware of and was labelled for both emotion and signs of emotion. Both the selection and the design of the labelling scheme were based on systematic criteria (see below) derived in part from the psychological literature on emotion and in part from experience with real data. The chapter "Issues in Data Labelling" in Part III gives further information on the principles and models which underpin the labelling scheme.

3.2 Size and Structure of the Labelled Subset

The labelled subset consists of 48 clips (between 3 s and 2 min in length) selected from the primary recordings. The labelled episodes are mounted on the ANVIL platform. Table 4 provides a summary of the numbers of clips selected from the

Table 4 Selection of clips for labelled subset

Raw data records	Data type	Number of clips selected
Belfast Naturalistic Database	Naturalistic	10
Castaway Reality Television Data Set	Naturalistic	10
Sensitive Artificial Listener (Belfast recordings in English)	Induced	12
Sensitive Artificial Listener (Tel Aviv recordings in Hebrew)	Induced	1
Activity Data/Spaghetti Data	Induced	7
Green Persuasive Data Set	Induced	4
EmoTABOO	Induced	2
DRIVAWORK (driving under varying workload) Corpus in German	Induced	1
GEMEP Corpus	Acted	1

range of data types to make up the labelled subset. Appendix 3 describes in detail each clip selected for labelling, including a summary of the content of the clip and descriptors of the emotion, context and modalities.

The selection of the final sample involves non-trivial issues. Two levels were used.

The first was the selection of clips from within a whole recording. In the case of relatively intense emotional episodes, the extraction of a section/clip includes build-up to and movement away from an emotional nucleus/explosion – lead in and coda are part of identification of the state. In the case of less emotionally intense recordings, the basic criterion used to set the boundaries of clips is that 'the emotional ratings based on the clip alone should be as good as ratings based on the maximum recording available' (i.e. editing should not exclude information that is relevant to identifying the state involved).

The second stage of selection was deciding which clips should form the labelled subset of the HUMAINE Database. It is very easy to drift into using a single type of material which conceals how diverse emotion actually is. To counter that, 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.

Table 5 shows the framework which underpinned the final selection. Each clip was chosen for its representation in those main classes.

The details of each of the 48 clips are given in Appendix 3. For each clip there is a short description of what is happening in the clip and the gender of the speaker as well as a description of quadrant, emotion mix, modality and context (constraints, goal and setting) as set out in Table 5. The clips are all available under Conditions of Use Agreement (Appendix 1).

The grid of distinctions underpinning the selection of clips is in some senses what one might expect – especially attention to emotional range, context and modality. But the history of emotional databases shows that this is not the norm. Very often

Table 5 Grid of distinctions underpinning selection of clips in labelled subset

Quadrant		Positive active, pragmatic, negative passive, negative active
Emotion mix	Within a clip	Consistent, co-existent (with another emotion) Shift (from one emotion to another)
Gender		Male/female
Modality		Face, speech, gesture
Context labels	Constraints	Un/restrained, Un/constrained
	Goal	Inform, create rapport, pure expression of emotion
	Setting	Task, activity, interactive, group, interactive-passive, other

databases are application focused and so they only have one type of data or where they are more open ended, they often do not prespecify criteria for selection from the raw records. This can lead to a slanted representation of the data. There are also some aspects of the grid that have not traditionally formed part of emotion databases. Specifying context is often not done, although the impact of context on the representation of an emotion can be large. And selection which takes into account the consistency of emotion and whether it is pure or co-existent with another emotion is a departure which is based on experience with real data. Experience with the Belfast Naturalistic Database and the EmoTV Database suggests that emotion mixing and emotion shifting are common (Devillers et al., 2006). Hence the grid is designed to capture clips which represent this feature.

3.3 Labelling of the Subset

A wide range of emotion labels and signs of emotion descriptors are attached to each clip. These are also available together with labelling manuals on the portal at www.emotion-research.net. The labelled data can be displayed on the ANVIL platform and the procedure for obtaining and using the software is explained on the portal. The emotion labelling has been done by six raters for all 48 clips and the data for all six labellers is available via the portal. The speech and language labelling has been done by one trained phonetician and is available for all 48 clips. Two clips are labelled for face and gesture; the labelling is available on the portal.

The components for the labelling scheme and the principles behind it are described in full in the chapter "Issues in Data Labelling". This section summarises how they have been put together and used in the labelled subset of the HUMAINE database.

3.4 Emotion Labels

Two levels of description are included.

At the first level, global labels are applied to an emotion episode or clip as a whole. Factors that do not vary rapidly (the person concerned, the context) are described here. This provides an index that can be used to identify clips that a particular user might want to consider. For instance, it will allow a user to find examples of the way anger is expressed in relatively formal interactions (which will not be the same as the way it is expressed on the football terraces).

Labelling at the second level is time aligned. This is done using 'trace'-type programs (see the chapter "Issues in Data Labelling"). Each of the programs deals with a single aspect of emotion (e.g. its valence, its intensity, its genuineness). An observer traces his/her impression of that aspect continuously on a one-dimensional scale while he or she watches the clip being rated. The data from these programs is

imported into ANVIL as a series of continuous time-aligned traces. The trace-type labelling captures perceived flow of emotion.

Tables 6 and 7 summarise the global and continuous 'trace' labels that are applied to each clip.

3.5 Sign Labels

Labels for signs of emotion are also attached to each clip in the labelled subset. There are labels for speech and language applied to all clips and labels for gesture and face applied to two of the clips. Table 8 summarises the descriptors used.

3.6 The Labelled Subset: Illustrations and Issues

The labelled subset is a powerful demonstration of the range and diversity of emotional life and how we can begin to describe it. This section works through a number of examples to illustrate some of the diversity and complexity of data in the subset and to show how the labelling can capture what is going on in quite complex emotional episodes.

Example 1. For the first example we return to the Spaghetti Data and to the woman already featured in Fig. 5. Table 9 shows how one rater described this clip in global emotional terms. Figure 9 shows the form of the display of emotional Trace continuous labelling for the same clip by the same rater. In Fig. 10 we see the whole of the clip labelled using the Trace programs. The screenshot is taken from ANVIL and it shows the traces from one rater for this clip for intensity of emotion, acting, masking, activation and power/powerlessness. The screenshot conveys the net effect of putting traces together. The clip shows a participant feeling in a box and suddenly triggering a buzzer. She gives a gasp, then a linguistic exclamation. The top trace, emotional intensity, rises abruptly after the gasp. The rater does not judge that the response is acted, but there is a degree of masking at the beginning which breaks down abruptly at the unexpected event. Activation rises abruptly after a delay (during which the participant might be described as frozen).

Example 2. The second example illustrates how emotion can look in different contexts and time domains. All six raters attached the word 'fear' to the clips from which the examples below are taken (Fig. 10). The first clip shows a subject we have seen before watching a friend fall off a mountain bike in the Belfast Activity Data. The second shows a subject undertaking the Spaghetti task at the point at which she stopped the task and said she was too frightened to continue. The third clip shows a subject recalling touching snakes in a darkened hut a few minutes after the event. The fourth shows a subject recalling a terrifying incident a year after it happened. These clips suggest that a sensible database needs to show the variety of things that are called 'fear'. The sample in the HUMAINE database is by no means complete, but it is a useful pointer to the variety of representations behind an emotion word.

Table 6 Global emotion descriptors applied to HUMAINE database (open comment is also invited for each class)

Classes of global emotion label	Description
Emotion words	Choice of up to 6 from list of 48
Emotion-related states	Choice from one or more of the following: attitudes, established emotion, emergent emotion (full blown), emergent emotion (suppressed), moods, partial emotion (topic shifting), partial emotion (simmering), stance towards a person, stance towards an object or a situation, interpersonal bonds, altered state of arousal, altered state of control, altered state of seriousness
Combination types	Choice from one or more of the following: unmixed emotion, simultaneous combination, sequential combination
Authenticity	Choice from a six-point scale, describing (i) degree of acting, (ii) degree of masking
Core affect dimensions	Choice from a six-point scale describing perceived intensity, activation and valence (0–6 for intensity and activation; –3 to +3 for valence)
Basic factual information	Choice from selection of labels describing age, gender and nationality of the person observed, the quality of the recording, the degree to which the person observed is physically constrained and the type of audience (if appropriate) present in the clip (close, colleague, public, artificial)
Context	Choice of labels describing the context of the communication taking place. These include the purpose of the communication (to inform/persuade/create rapport/destroy rapport/pure expression), social setting (none, passive other, interactant, group) and social pressure (to formality, weak, expressiveness). The rater is also asked to judge the degree of camera and microphone awareness
Key events/emotional focus	Up to three descriptions of the emotional focus of the clip. This includes identifying the emotional focus of the clip and selecting a qualifying temporal label – current, recalled, anticipated or imagined
Other key events	A description of other key events which may be relevant to creating the emotions present in the clip but are not themselves the emotional focus of the clip. The key event is further qualified by identifying whether it is a trigger (short term), cause (long term) or aspiration (future)
Appraisal categories	A rating of how strongly the observed person's emotional state is related to aspects of the way he/she sees the emotionally significant events or people around. The factors are goal conduciveness and goal obstructiveness; power/powerlessness; expectedness

Table 7 'Trace' label descriptors

Trace label	Description
IntensTrace	A rating of the intensity of the emotion that the specified person is perceived to be experiencing from moment to moment. The range of intensity runs from the person experiencing no emotion whatsoever, i.e. as emotionally still as they could be, to their emotion being perceived as being as intense as it could possibly be
ActTrace	A rating of the extent to which the specified person is trying to give an impression of emotions that they actually do not feel (i.e. they are pretending or acting). The range runs from no attempt to simulate unfelt emotions, i.e., they would be perceived as being totally genuine in their emotional expression, to where their emotion would be perceived as completely acted or false
MaskTrace	A rating of the extent to which the specified person is trying to avoid showing emotions that they do actually feel (i.e. the extent to which they are trying to cover up their genuine emotion). The range runs from no attempt to avoid showing their emotions, i.e. they would be perceived as being totally open in their emotional expression, to where their emotion would be perceived as completely masked or covered
ActivTrace	A rating of activation or arousal, i.e. how strongly the relevant person is inclined to take action. The range runs from 'absolutely no inclination to be active' through markers 'weakly active', strongly active, to 'compelling urge to be active' at the maximum of the scale
ValenceTrace	A rating of how positive or negative the specified person feels about the events or people at the focus of his or her emotional state. The range runs from very strongly negative to very strongly positive. At the mid-range, they would be perceived as being neutral or experiencing no emotion
PowerTrace	A rating of how in control the specified person feels about the events or people at the focus of his or her emotional state. The range runs from 'absolutely no control over events', through 'not quite in control', 'just about in control' to being completely in control over events at the maximum of the scale
Anticipate/ ExpectTrace	A rating of the extent to which the specified person has been taken unawares by the events at the focus of their emotional state. The range runs from events being anticipated completely to being taken completely unawares by the event
WordTrace	A rating of the intensity of the four highest ranked emotion words for that clip The range runs from absolutely none of that emotion being present to that emotion being expressed as purely as it could be

Table 8 Signs of emotion descriptors

Speech and language descriptors	Gesture descriptors	Face descriptors
Transliteration (*words spoken*)	Gesture units	FAPS (automatically derived)
Largely automatically derived labels *(time waveform, pitch)*	Phases	
Auditory-based labels (*paralanguage, voice quality, timing, volume*)	Phrases/categories Lemmas adapted from gesture lexicon	

Example 3. The third example focuses on portraying the complexity of emotions that can occur in naturalistic data, particularly the mixed nature of emotions (referred to as 'co-existing' in the HUMAINE coding scheme) and the way in which emotion fluctuates and shifts within short time periods. By comparison, acted data tends to portray emotion as consistent, pure and static over a period of time. Figure 11 shows a sequence of frames from the Belfast Naturalistic Database taken at intervals from a 10-s period in which the subject utters the words 'It's a boy. And the anger drained out of me that night. I felt it going. It was like a release.'

The context for the sequence of frames is that the subject is remembering the birth of her grandson (her daughter's child). She and her daughter had not got on very well together and she had been particularly angry at her daughter for getting pregnant. She has been describing the anger she felt but then moves to describe the moment of her grandson's birth and the release from the anger she had been feeling when the moment she heard her grandson had been born. Figure 11 shows the way in which the emotion shifts and blends as she recalls the incident. The sequence is typical of the type of data that comes from naturalistic settings. Work on EmoTV (Devillers et al., 2006), which unfortunately cannot be released for copyright reasons, makes similar points. One of the interesting things about this particular example is that the emotion expressed comes from recalling events, illustrating that recall can produce fairly intense emotion.

Example 4. This is a nice illustration of the need to consider a wide range of emotion-related states when classifying emotional behaviour. The opening chapter of this handbook discusses the theory behind these, and Table 6 lists those that are used in labelling the HUMAINE database. Figure 12 illustrates one of these states which is less commonly talked about – 'suppressed' emotion. The subject is shown talking to the angry personality of the Sensitive Artificial Listener in the SAL data (see above). What is happening is that the angry personality of SAL is trying to wind up the subject's emotions into an angry state. All the raters of this clip attach the label 'suppressed emotion' to it. The words that they also all agree apply to the clip are politeness, tension, irritation, annoyance and anger. These are in line with the global label of 'suppressed emotion', indicating that the subject may have negative emotions but that he remains polite, keeping his emotions under control through a deliberate effort. The text (see under Fig. 12) indicates suppression of the emotions.

Table 9 Global emotion labels for Spaghetti Data Clip 14e (from one rater)

Emotion words (in order of relevance)	Anxiety 1 Shock 2 Surprise 3 Relieved 4
Emotion-related states	Stance towards object/situation Altered state of control Emergent emotion (full blown) Altered state of arousal
Combination types	Sequential combination Simultaneous combination (rater's comment: anxiety first, then a combination of shock and surprise, then relief)
Authenticity	Masking strong Acting moderate
Core affect dimensions	Intensity 5 Activation 5 Valence –1
Basic factual labels	Video quality: good Hands constrained: no Posture constrained: no Nationality: N Irish Recording style: observation, task Acoustic quality: good Physically unrestricted: yes Gender: F Age: 18–25 Type of audience: colleagues
Context	Social pressure: weak Camera and microphone awareness: yes Passive other Pure expression of emotion
Key events	What the left hand encounters in the box Here and now
Other key events	–
Appraisals	Goal obstructiveness: open (i.e. no judgement) Goal conduciveness: open (i.e. no judgement) Expectedness: unanticipated Power/powerlessness: very powerless

Examples 5 and 6 illustrate the labelling of signs of emotion and the richness of signs in the data. Figure 13 shows a frame from EmoTABOO with the array of gesture labels and FAPs attached. Figure 14 shows a frame from Castaway Reality TV. The episode is particularly rich in paralinguistic expression of emotion. The subject is asked what he misses most. He replies: ' I always get choked up . . . . . family' accompanied by long pausing, tremulous voice and nervous laughter.

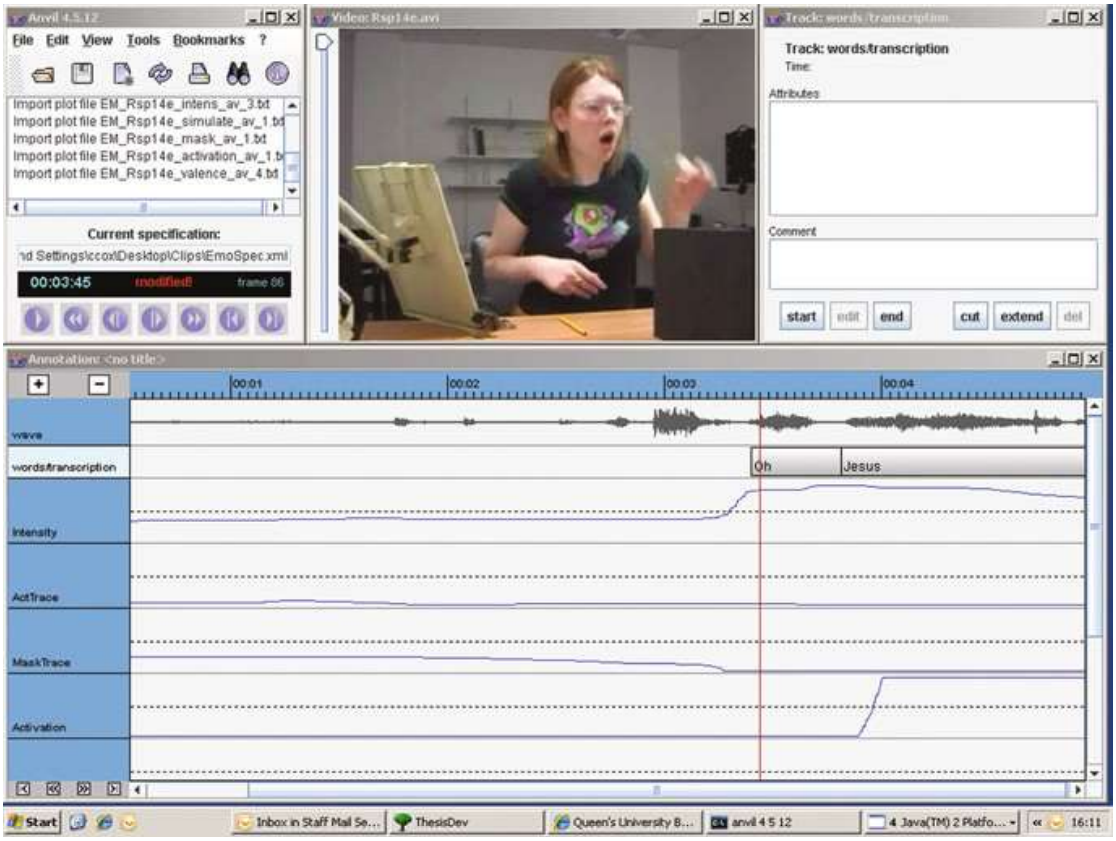

Fig. 9 Continuous trace labelling for intensity, acting, masking and level of activation for Clip 14e by one rater (*red vertical bar* marks frame shown)

Fig. 10 Four representations of fear

Fig. 11 Clip 56d Belfast Naturalistic Database. Frame (a) starts with the memory of the announcement by the daughter's birth partner that the baby had been born and that it was a boy. The expression certainly seems to contain happiness. In frames (b), (c) and (d) she recalls the anger she had felt but at the same time recalls her move away from the anger: the frames are clearly a mix of complex feelings (anger, pain, sadness, escape) and emotional shift. Frames (e) and (f) describe release from the anger and the final two frames might best be described as a return to peace

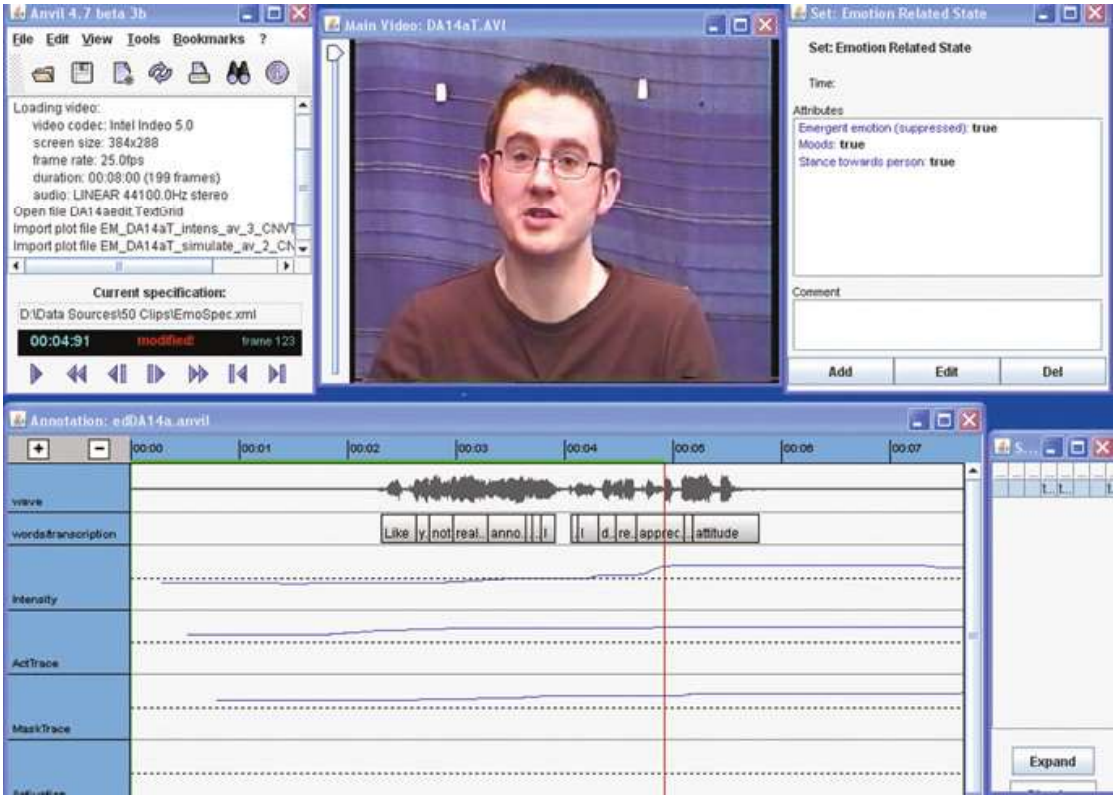

Fig. 12 An example of suppressed emotion from SAL Data as subject talks to the angry personality Spike. (Text: Like you're not really annoying me and I don't appreciate your attitude)

Fig. 13 Gesture labelling and FAPS applied to EmoTABOO Data

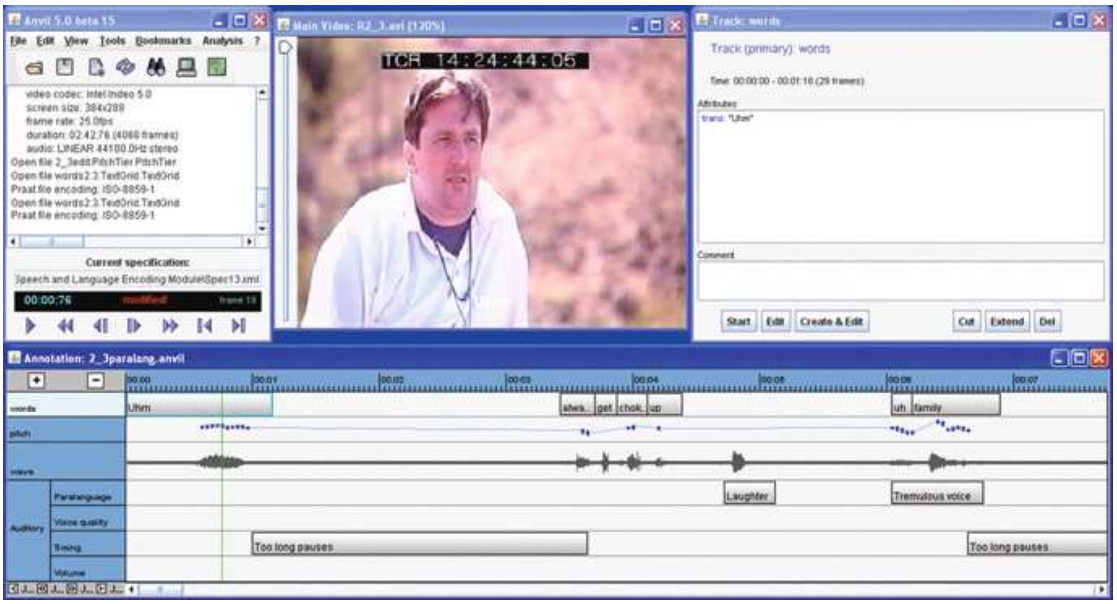

Fig. 14 Paralinguistic expression of emotion in Castaway Reality TV Data Set

Example 7. The final figure, Fig. 15, shows some of the interesting and unexpected interactions between modalities. The subject in it appears to be smiling but the words indicate that she is actually in a state of shock. The subject (from the Castaway Reality TV Data Set) is recalling her encounter with snakes in the hut from which she has just emerged. She is saying 'and the first thing I touched was the snake's head . . . feel really shaky now.' The bottom line in the figure is a trace

from WordTrace by one rater. The word that the rater thought applied most to the episode was ‘fear’ and the trace is a trace of fear in the episode. The trace shows that fear is present and strong throughout the episode. There are many examples in the HUMAINE database where the expression on the face seems at odds with the emotion experienced.

4 Future Directions

The point of a database is to facilitate research, and the HUMAINE Database opens up a very large number of avenues for exploration.

At the most routine, the database provides evidence on the way a range of tools function and therefore provides a basis for evaluating them. The tools include everyday emotion categories, trace programs, descriptors for broad types of emotion and context. Evaluations include simple, formal procedures, such as tests of reliability. However, they also include others which are less clear-cut, but not less important: does the battery of descriptions tell us what we need to know, and if not, why not?

Related, but distinct, are questions about reduction. There are two obvious forms of question to consider. The first form is related to the concept of cover classes. It is concerned with establishing which labels can be merged without unacceptable loss of information. The second form is related to the concept of dimensions. There is a very large literature on the number of dimensions needed to represent a set of words. The HUMAINE Database opens up the possibility of asking how many dimensions are necessary to represent a set of samples of emotionally coloured behaviour.

These questions are not statistically trivial. For example, they should ideally take account of the way labellings evolve over time. A simplification which seems fair in terms of a series of ‘snapshots’ may be a serious problem if it undercuts the ability to predict what will happen next. Standard statistical reduction techniques do not address that kind of problem.

The end target of that kind of work is an empirically validated set of labels. It is frustrating, but there is no way to reach that stage without generating labellings some of whose components will eventually be discarded. Hence, it is to be expected that some components of the HUMAINE scheme will be discarded in the process of analysis. Conversely, new components will presumably need to be added. Iterative adjustment is to be expected, but it needs a core to work round.

Benchmarking is another key application. It is a major problem that the area lacks standard tasks against which the performance of new algorithms can be tested. The database offers two kinds of benchmark – clips to be analysed and types of information to be recovered. The test is not confined to machine recognition. For instance, the material in the database offers a very interesting test for brain-scanning technologies. Capturing differences between responses to HUMAINE Database clips is a much more acute test than is capturing differences between responses to photographs from the standard Ekman collection.

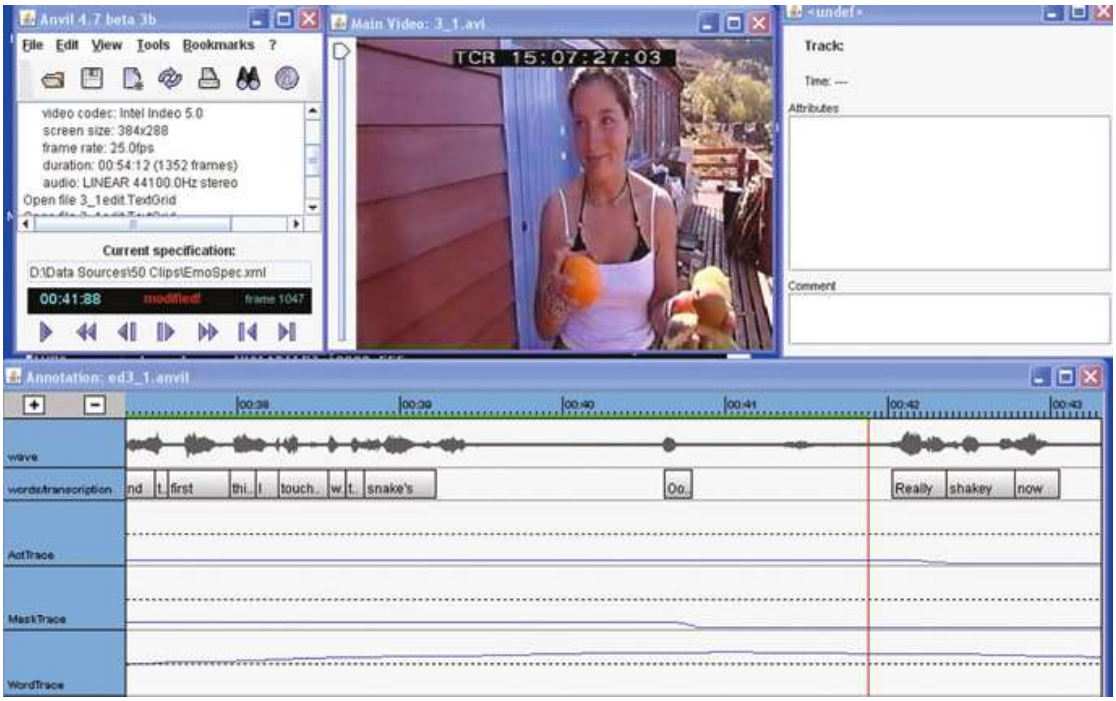

Fig. 15 Conflicting signs from face and speech, Castaway Reality TV Data Set

The records lend themselves to a range of studies. The most routine is simply extending the labelled set. There is a very large body of primary records that remains unlabelled, and its value would be multiplied if labelled versions were available to the community.

A wide range of issues call out for more specific studies. Three will be singled out here. The first is relationships between modalities. It has been pointed out that impressionistically, audio and visual signals sometimes seem to point in very different directions. The data provides opportunities to explore that issue much more systematically. A natural starting point is simply to label a substantial body of material on the basis of audio records only, visual records only and verbal transcripts only. The second is temporal evolution of expressions. Examples like Fig. 10 make it clear that there are rapid, radical changes in moment-by-moment expression of emotion. Some theoretical frameworks predict that change should occur on that kind of timescale (Scherer and Ellgring, 2007). The HUMAINE Database contains a reservoir of naturalistic data that makes it possible to explore these ideas. The third is whether information is localised or distributed. It is not obvious whether information about a person's emotion is, so to speak, smeared evenly over time or concentrated in a few revealing moments. The material in the database invites research on the topic.

Beyond these, the database provides a kernel of primary records that help to clarify which kinds of extension make sense. It certainly does not make sense to collect new records at random. A considerable range of states and contexts are probably quite well covered in the data that now exists, and random inventions are quite likely to produce nothing but more stilted variations on the same theme. However, there

are areas where information is quite clearly limited, and those are the areas where it makes sense to concentrate effort. An overwhelmingly obvious example is cross-cultural difference. A few HUMAINE techniques have been applied in substantially different cultures. The process needs to be extended radically to provide anything approaching a reasonable representation of the way culture affects the expression of emotion.

Perhaps most fundamental of all, the database invites a cumulative and collaborative attitude. The HUMAINE Database is a product of collaborations between several teams over a period of years. It will take a much larger scale of collaboration to accumulate a reservoir of data sufficient to understand the various ways in which various kinds and combinations of emotion can colour various actions and interactions.

Appendix 1: User Agreement – HUMAINE Database

This user agreement applies to the database *HUMAINE database* provided on the enclosed CDs. It has been released by QUB (Queens University Belfast) under specific conditions for sole scientific, non-commercial use.

Conditions of Release

- Data set provided without guarantee.
- No legal claims of any kind can be derived from accepting and using the data set.
- QUB is not liable for any damage resulting from receiving, installing or using the data set or other files provided by QUB in this context.
- Expressed written consent must be sought from QUB before the data set or any other files provided by QUB containing information derived from it (e.g. labelling files) are passed on by the licensee to any third party.
- If a partner/user concentrates on labelling, s/he agrees to share with all the other HUMAINE partners additional analyses, especially additional annotations. In the first instance, the analyses/annotations should be returned to the HUMAINE team at QUB.
- Any models derived using data from the data set may be used only for scientific, non-commercial applications.
- The licensee will let QUB know of any results without undue delay. Joint publications (between the licensee and QUB) should be aimed at.

For publications concerning direct or indirect use of the corpus, the licensee must cite QUB with a citation provided by QUB.

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

Signed_____

Please print name_____

Appendix 2

Data source and reference	Nature of material	Emotional content	Technical info and availability
Belfast Naturalistic Database (Douglas-Cowie et al., 2003)	The material consists of audiovisual sedentary interactions from TV chat shows and religious programs, and discussions between old acquaintances	A range of positive and negative emotions. Intensity is mostly moderate	There are 125 subjects (two sequences of 10–60 s each, one neutral and one emotional); each sequence has attached to it three emotion words which best describe it attributed by three listeners from a list of some 50 possible words; a selection of 30 sequences with ethical and copyright clearance is available
EmoTV Database (in French) (Devillers et al., 2006)	The EmoTV database consists of audiovisual interactions from TV interviews – both sedentary interactions and interviews 'on the street' (with wide range of body postures)	A range of positive and negative emotions. Intensity moderate with some intense material	48 subjects (51 sequences of 4–43 s per subject in emotional state); copyright restrictions prevent straight release, but there may be circumstances under which some data can be shared
Castaway Reality Television Database (Douglas-Cowie et al., 2007)	This consists of audiovisual recordings of a group of 10 taking part competitively in a range of testing activities (feeling snakes, lighting outdoor fires) on a remote island. The recordings include single and collective recordings and post-activity interviews and diary-type extracts	A range of positive and negative emotions. Intensity moderate with some intense material	10 tapes of 30 min each; copyright clearance

(continued)

Data source and reference	Nature of material	Emotional content	Technical info and availability
Sensitive Artificial Listener (Douglas-Cowie et al., 2007)	The SAL data consists of audiovisual recordings of human–computer conversations elicited through a 'Sensitive Artificial Listener' interface designed to let users work through a range of emotional states (like an emotional gym). The interface is built around four personalities – Poppy (who is happy), Obadiah (who is gloomy), Spike (who is angry) and Prudence (who is pragmatic). The user chooses which he/she wants to talk to. Each has a set of stock responses which match the particular personality. The idea is that Poppy/Spike/Obadiah/Prudence draws the user into their own emotional state	A wide range of emotions but they are not very intense	Data has been collected for four users with around 20 min of speech each. SAL has also been translated into Hebrew (at Tel Aviv University) and Greek (at National Technical University of Athens, ICCS) and adjusted to suit cultural norms and expectations, and some initial data has been collected. The data has ethical permission and is available to the research community

(continued)

Data source and reference	Nature of material	Emotional content	Technical info and availability
Activity Data/Spaghetti Data (Douglas-Cowie et al., 2007)	Audiovisual recordings of emotion in action were collected using two induction techniques developed in Belfast. In the first, volunteers were recorded engaging in outdoor activities (e.g. mountain bike racing). The second used a more controlled environment where certain kinds of 'ground truth' could be established. It is called the Spaghetti method, because participants are asked to feel in boxes in which there were objects including spaghetti and buzzers that went off as they felt around. They recorded what they felt emotionally during the activity	Method 1 elicited both positive and negative emotions with a high level of activation. Method 2 elicited a range of brief, relatively intense emotions – surprise, anticipation, curiosity, shock, fear, disgust	Method 1 produced 'provocative' data which was very fast moving and had a noisy sound track. Method 2 produced data where the participants were reasonably static and stayed within fixed camera range, making it easier to deal with face detection. The audio output consists mainly of exclamations. There are now recordings of some 60 subjects. The data has ethical permission and is available to the research community

(continued)

Data source and reference	Nature of material	Emotional content	Technical info and availability
Belfast Driving Simulator Data (Douglas-Cowie et al., 2007) http://emotion-research.net/ws/wp5/edelle.ppt	The driving simulator procedure consists of inducing subjects into a range of emotional states and then getting them to drive a variety of 'routes' designed to expose possible effects of emotion. Induction involves novel techniques designed to induce emotions robust enough to last through driving sessions lasting tens of minutes. Standard techniques are used to establish a basic mood, which is reinforced by discussions of topics that the participants have preidentified as emotive for them. The primary data is a record of the actions taken in the course of a driving session, coupled with physiological measures (ECG, GSR, skin temperature, breathing). It is supplemented by periodic self-ratings of emotional state	Three emotion-related conditions, neutral, angry, and elated	30 participants; will be available pending completion of Ph.D. on the data

(continued)

Data source and reference	Nature of material	Emotional content	Technical info and availability
EmoTABOO (in French) (Zara et al., 2007)	EmoTABOO records multimodal interactions between two people during a game called Taboo. One person has to explain to the other using gestures and body movement, a 'taboo' concept or word	Range of emotions including embarrassment, amusement	By arrangement with the LIMSI team
Green Persuasive Data Set	The data set consists of audiovisual recordings of interactions where one person tries to persuade another on a topic with multiple emotional overtones (adopting a 'green' lifestyle)	Complex emotions linked to varied cognitive states and interpersonal signals	Eight interactions of about 30 min each, and associated traces made by the interviewees to indicate how persuaded they felt from moment to moment. The data has ethical permission and is available to the research community
DRIVAWORK (driving under varying workload) corpus (Honig, 2007)	The DRIVAWORK corpus has been collected at Erlangen, using a simulated driving task. There are three types of episode: participants are recorded relaxing, driving normally or driving with an additional task (mental arithmetic). Recordings are video and physiological (ECG, GSR, skin temperature, breathing, EMG and BVP)	Stress-related states rather than emotion per se	Availability by arrangement with Erlangen team

Appendix 3

Clip/source	Content	Gender	Emotion words	Quadrant	Combination type	Modalities present (*=labelled)	Context	Goal	Constraints
APal/Emo Taboo		M/M	Amusement, empathy, frustration, friendliness, embarrass-ment, powerlessness			Face*, speech*, gesture*	Interactive task	Inform	Unrestricted, unconstrained
Cpal/Emo Taboo		M/F	Anxiety, amusement, embarrass-ment, friendliness interest, politeness			Face*, speech*, gesture*	Interactive task	Inform	Unrestricted, unconstrained
2_3/Castaway Reality TV Data Set	Subject describes how he misses his family and how much he has learnt about himself from experiences on island	M	Helplessness, love, power-lessness, satisfaction, stress, worry	NegPass	Consistent emotion	Face, speech*	Passive other	Inform	Unrestricted, constrained

(continued)

Clip/source	Content	Gender	Emotion words	Quadrant	Combination type	Modalities present (*=labelled)	Context	Goal	Constraints
4_2/Castaway Reality TV Data Set	Subject describes his fear after the insect task and says he does not want to continue	M	Stress, anxiety, doubt, shame, relieved, fear	NegPass	Consistent emotion	Face, speech*	Passive other	Inform	Unrestricted, constrained
10_1/Castaway Reality TV Data Set	Subject, who has decided to leave, says goodbye to fellow castaways and makes plans to meet up in the future	M	Friendliness, relieved, helplessness, affection, politeness, powerlessness	Pos Act	Consistent emotion	Face, speech*, gesture	Group	Rapport	Unrestricted, unconstrained
10_3/Castaway Reality TV Data Set	Group leader tells group they can have a shower to which they respond by cheering, laughing, smiling	Group M/F	Relaxed, interest, calm, happiness, politeness, tension	Pos Act	Shifting emotion	Face, gesture	Group	Inform	Unrestricted, unconstrained
R56b/Belfast Naturalistic Database	Subject describes her love for her son-in-law	F	Love, satisfaction, affection, pride, friendliness, happiness	Pos Pass	Consistent emotion	Face, speech*	Passive other	Inform	Unrestricted, constrained

(continued)

Clip/source	Content	Gender	Emotion words	Quadrant	Combination type	Modalities present (*=labelled)	Context	Goal	Constraints
R56d/Belfast Naturalistic Database	Subject describes the happiness and excitement at the birth of her grandchild	F	Happiness, joy, relieved, affection, politeness, satisfaction	Pos Act	Co-existing emotion	Face, speech*	Passive other interactive	Inform	Unrestricted, constrained
1_4/Castaway Reality TV Data Set	Subject describes his experiences, e.g. an argument with a fellow contestant and his lowest point of the experience	M	Annoyance, amusement, relaxed, satisfaction, friendliness, frustration	Pos Act	Co-existing + shift	Face, speech*	Interactive	Inform/ activity	Unrestricted, unconstrained
3_1/Castaway Reality TV Data Set	Subject comes out of house in which she touched insects and snakes, and speaks of her fear and disgust	F	Relieved, anxiety, fear, disgust, amusement, powerlessness	NegAct – P	Co-existing emotion	Face, speech*, gesture	Interactive	Inform/ activity	Unrestricted, unconstrained

(continued)

Clip/source	Content	Gender	Emotion words	Quadrant	Combination type	Modalities present (*=labelled)	Context	Goal	Constraints
5_2/Castaway Reality TV Data Set	Subject talks about his feeling of surprise when the other contestants voted for him and how he has changed and developed during the experience	M	Satisfaction, pride, happiness, tension, content, amusement	NegPass	Consistent emotion	Face, speech*, gesture	Interactive	Inform/ activity	Unrestricted, uncon-strained
5_3/Castaway Reality TV Data Set	Subject tells her theory that seven contestants have already been chosen by the organisers to leave, despite the contestants having been told that they could choose themselves	F	Amusement, politeness, powerless-ness, contempt, annoyance, interest	Prag+P Act	Consistent emotion	Face, speech*	Passive other interactive	Inform	Unrestricted, uncon-strained
5_3/Castaway Reality TV Data Set	Group discusses various things – focus is on subject's facial expression	M	Boredom, irritation, calm, politeness, worry, amusement	NegPass	Consistent emotion	Face, gesture	Passive other	Pure expres-sion	Unrestricted, uncon-strained

(continued)

Clip/source	Content	Gender	Emotion words	Quadrant	Combination type	Modalities present (*=labelled)	Context	Goal	Constraints
056e/Belfast Naturalistic Database	Subject speaks of being attacked on leaving her house and describes fear at thinking husband had been shot when gun was fired	F	Fear, calm, anxiety, tension, powerlessness sadness	NegPass	Consistent emotion	Face, speech*	Interactive	Inform	Unrestricted, uncon-strained
077a/Belfast Naturalistic Database	Subject describes the beautiful setting of friends' home in which the bedroom overlooks a golf course	F	Interest, pleasure, friendliness, pride, calm, relaxed	Prag	Consistent emotion	Face, speech*, gesture	Interactive	Inform	Unrestricted, uncon-strained
077c/Belfast Naturalistic Database	Subject describes pain and emotional trauma following an accident	F	Sadness, despair, hurt, anxiety, calm, worry	NegPass	Consistent emotion	Face, speech*	Interactive	Inform	Unrestricted, uncon-strained
078a/Belfast Naturalistic Database	Subject speaks of working as a diesel fitter in the company managed by his father	M	Interest, pride, relaxed, friendliness, politeness, annoyance	Prag	Consistent emotion	Face, speech*, gesture	Interactive	Inform	Unrestricted, uncon-strained

(continued)

Clip/source	Content	Gender	Emotion words	Quadrant	Combination type	Modalities present (*=labelled)	Context	Goal	Constraints
079b/Belfast Naturalistic Database	Subject describes her amusement at grandmother thinking her father would have to come home from overseas to help with a medical situation in his own country	F	Amusement, affection, friendliness, happiness, sadness, love	Pos Act	Consistent emotion	Face, speech*, gesture	Passive other interactive	Inform	Unrestricted, uncon-strained
R56g/Belfast Naturalistic Database	Subject describes determination to continue her job, despite being threatened	F	Annoyance, irritation, anger, frustration, pride calm	NegAct	Consistent emotion	Face, speech*	Passive other interactive	Inform	Unrestricted, constrained
R77d/Belfast Naturalistic Database	Subject describes grandson, who has just begun to walk, and how his mother treats him as a younger baby by mashing his food	F	Happiness, love, affection, pride, excitement, joy	Pos Act	Consistent emotion	Face, speech*, gesture	Passive other interactive	Inform	Unrestricted, uncon-strained

(continued)

Clip/source	Content	Gender	Emotion words	Quadrant	Combination type	Modalities present (*=labelled)	Context	Goal	Constraints
R78b/Belfast Naturalistic Database	Subject describes anger at the fact that a politician is preventing his father from voting when his father fought for the country and the politician did not	M	Annoyance, frustration, irritation, contempt, calm, anger	NegAct	Consistent emotion	Face, speech*, gesture	Passive other interactive	Inform	Unrestricted, uncon-strained
RCsh/Belfast Outdoor Activity Data	Subject's scream-ing/laughing response to a bicycle crash	F	Excitement, amusement, pleasure, happiness, elation, shock	PA/NA Mix	Co-existing emotion	Face, gesture	Activity	Pure expres-sion	Unrestricted, uncon-strained
REdA1/ Sensitive Artificial Listener Data Set	Subject speaks to character about a fantasy holiday on a desert island	F	Amusement, happiness, pleasure, friendliness, embarrass-ment, joy	Pos Act	Consistent emotion	Face*, speech*	Interactive	Inform/ Rapport	Restricted, constrained
REllA2 Sensitive Artificial Listener Data Set	Subject speaks to gloomy character Obadiah, who offers her no help or advice	F	Irritation, annoyance, contempt, tension, frustration, satisfaction	PA/NA Mix	Shifting emotion	Face, speech*	Interactive	Inform	Restricted, constrained

(continued)

Clip/source	Content	Gender	Emotion words	Quadrant	Combination type	Modalities present (*=labelled)	Context	Goal	Constraints
RElIB2 Sensitive Artificial Listener Data Set	Subject says she would like to go for a cycle on a summer evening, describing the hawthorn she would smell and how the air would blow around her	F	Pleasure, happiness, delight, interest, politeness, relaxed	Prg/NA Mix	Consistent emotion	Face, speech*	Interactive	Inform	Restricted, constrained
Rsp14e/ Spaghetti	Subject puts hand in box, which she cannot see inside, and is startled by a noise from the box	F	Shock, anxiety, surprise, amusement, fear, relieved	NA/NP Mix	Shifting emotion	Face, speech*, gesture	Passive other	Task	Unrestricted, constrained
8_3/Castaway Reality TV Data Set	Subject tries to light a fire, which she finds difficult, but eventually succeeds	F	Frustration, annoyance, amusement, irritation, satisfaction, tension	Pos Act	Co-existing emotion	Face, speech*, gesture	Task	Activity	Unrestricted, uncon-strained
AO4joi/Geneva Acted Emotion Database		M	Excitement, delight, elation, joy, happiness, pleasure	Pos Act	Consistent emotion	Face, gesture	Activity	Pure expres-sion	Unrestricted, uncon-strained

(continued)

Clip/source	Content	Gender	Emotion words	Quadrant	Combination type	Modalities present (*=labelled)	Context	Goal	Constraints
A16aT/Sensitive Artificial Listener Data Set	Subject says she finds it preferable to be open-minded, as more opportunities arise	F	Amusement, calm, courage, friendliness, relieved, embarrassment	Prag	Co-existing emotion	Face, speech*	Passive other interactive	Inform	Unrestricted, uncon-strained
Arithmetic/ DRIVAWORK Erlangen Campus		M	Amusement, interest, calm, irritation, frustration, stress	Prag	Shifting emotion	Face	Task	Activity	Unrestricted, constrained
B16aT/ Sensitive Artificial Listener Data Set	Subject claims that she does not get worked up about small things but is not a doormat	F	Annoyance, pride, calm, frustration, irritation, satisfaction	Neg P-A	Shifting emotion	Face, speech*	Passive other interactive	Inform	Unrestricted, uncon-strained
DA14AT/ Sensitive Artificial Listener Data Set	Subject tells character he doesn't like his attitude	M	Annoyance, irritation, politeness, tension amusement, boredom	NegAct	Consistent emotion	Face, speech*	Passive other interactive	Inform	Unrestricted, uncon-strained

(continued)

Clip/source	Content	Gender	Emotion words	Quadrant	Combination type	Modalities present (*=labelled)	Context	Goal	Constraints
DA19T/ Sensitive Artificial Listener Data Set	Subject hopes that they are not causing any suffering, and that he can contribute to relieving it	M	Calm, amusement, affection, friendliness, pleasure, surprise	NegPass-Prag	Co-existing emotion	Face, speech*	Interactive	Inform	Unrestricted, uncon-strained
EllB3/ Sensitive Artificial Listener Data Set	Subject tells Poppy that she knows how to smile and that she always looks happy because of her mouth and teeth	F	Happiness, amusement, affection, friendliness, pleasure, surprise	Pos Act	Consistent emotion	Face, speech*	Interactive	Inform	Unrestricted, constrained
Ex2A/Green Persuasion Data	Subject tells female listener of the advantages of using public transport and describes how eco-friendly cars are not always efficient	M	Interest, calm, relaxed, worry, hope, politeness	Prag-Pos Act	Shifting emotion	Face, speech*, gesture	Interactive	Persuade	Unrestricted, uncon-strained
Ex6A/Green Persuasion Data	Subject describes the potentially lethal effects of global warming to male listener	M	Interest, worry, calm, friendliness, frustration, stress	Prag-PA	Consistent emotion	Face, speech*, gesture	Interactive	Persuade	Unrestricted, uncon-strained

(continued)

Clip/source	Content	Gender	Emotion words	Quadrant	Combination type	Modalities present (*=labelled)	Context	Goal	Constraints
HSAL/Hebrew Sensitive Artificial Listener		F	Boredom, serene, content, calm, annoyance, anxiety	Prag	Consistent emotion	Face, speech, gesture	Passive other interactive	Inform	Unrestricted, constrained
P6aT/Sensitive Artificial Listener Data Set	Subject tells character how he tries to be happy with the present situation and hopes things will move forward in the future	M	Disappointment, worry, frustration, hope, calm content	NegPass	Consistent emotion	Face, speech*	Passive other interactive	Inform	Unrestricted, uncon-strained
PT2a/Green Persuasion Data	Shows responses of subject as she listens to the advantages of using public transport	F	Interest, politeness, boredom, amusement, calm, embarrassment	Pos Act-Prag	Shifting emotion	Face, speech*, gesture	Interactive	Persuade	Unrestricted, uncon-strained
PT6a/Green Persuasion Data	Subject listens to potential effects of global warming	M	Interest, politeness, boredom, doubt, relaxed, calm	Prag	Consistent emotion	Face, speech*, gesture	Interactive	Persuade	Unrestricted, uncon-strained

(continued)

Clip/source	Content	Gender	Emotion words	Quadrant	Combination type	Modalities present (*=labelled)	Context	Goal	Constraints
R12a/Sensitive Artificial Listener Data Set	Subject describes how feeling a wide range of emotions is a positive experience	M	Sadness, calm, frustration, interest, stress, relaxed	Neg P-Prag	Consistent emotion	Face, speech*	Passive other interactive	Inform	Unrestricted, unre-strained
RA21/Sensitive Artificial Listener Data Set	Subject tells character how simple things, like getting a good night's sleep, can improve negative emotions	M	Sadness, calm, frustration, hope, doubt, relaxed	Neg P-Prag	Shifting emotion	Face, speech*	Passive other interactive	Inform	Unrestricted, constrained
RB31t/ Sensitive Artificial Listener Data Set	Subject explains how rationality can be a stupid state to get into if things are going badly	M	Annoyance, irritation, anger contempt, frustration, excitement	NegAct	Consistent emotion	Face, speech*	Passive other interactive	Inform	Unrestricted, constrained
RhnT/ Belfast Outdoor Activity Data	Subject refuses to repeat task, which she failed to complete	F	Sadness, amusement, despair, helplessness, worry, disap-pointment	NegA	Co-existing emotion	Face, gesture	Task	Activity	Unrestricted, unre-strained

(continued)

Clip/source	Content	Gender	Emotion words	Quadrant	Combination type	Modalities present (*=labelled)	Context	Goal	Constraints
spag03/ Spaghetti	Subject feels in box and is disgusted at what she finds	F	Disgust, anxiety, amusement, fear, surprise, irritation	NegA-PA	Shifting emotion	Face, speech*, gesture	Task	Activity	Unrestricted, uncon-strained
spag11d/ Spaghetti	Subject is shocked at what she feels in box	F	Amusement, surprise, fear, anxiety, worry, tension	PosA-NP	Shifting emotion	Face, speech, gesture	Task	Activity	Unrestricted, uncon-strained
spag16d/Spaghetti	Subject is too frightened to feel deep into box and refuses to complete the task		Fear, anxiety, worry, doubt, tension, disgust	PosA-NP	Shifting emotion	Face, speech*, gesture	Task	Activity	Unrestricted, uncon-strained
WideEyeT/Belfast Outdoor Activity Data		F	Amusement, excitement, fear, surprise, worry, delight	NegA-PA	Shifting emotion	Face, gesture	Task	Activity	

References

Bänziger T, Scherer K (2007) Using actor portrayal to systematically study multimodal emotional expression: the GEMEP corpus. In: Paiva A, Prada R, Picard R (eds) Affective computing and intelligent interaction, Lisbon, Sept 2007. Springer LNCS , Berlin, pp 476–487

Cowie R, Douglas-Cowie E, Tsapatsoulis N, Votsis G, Kollias S, Fellenz W, Taylor J (2001) Emotion recognition in human-computer interaction. IEEE Signal Process Mag 32–80, January 2001

Devillers L, Cowie R, Martin J-C, Douglas-Cowie E, Abrilian S, McRorie M (2006) Real-life emotions in French and English TV video corpus clips: an integrated annotation protocol combining continuous and discrete approaches. In: LREC 2006, Genoa. http://www.cs.brandeis.edu~marc/misc/proceedings/lrec-2006/workshops/W09/Emotion-proceeding.pdf

Douglas-Cowie E, Campbell N, Cowie R, Roach P (2003) Emotional speech: towards a new generation of databases. Speech Commun 40(1–2):33–60

Douglas-Cowie E, Cowie R, Sneddon I, Cox C, Lowry O, McRorie M, Martin J-C, Devillers L, Abrilian S, Batliner A, Amir N, Karpouzis K (2007) The HUMAINE database: addressing the collection and annotation of naturalistic and induced emotional data. In: Paiva A, Prada R Picard R (eds) Affective computing and intelligent interaction, Lisbon, Sept 2007. Springer LNCS, Berlin, pp 488–500

Douglas-Cowie E, Devillers L, Martin J-C, Cowie R, Savvidou S, Abrilian S, Cox C (2005) Multimodal databases of everyday emotion: facing up to complexity. In: 9th European conference on speech communication and technology (Interspeech' 2005), Lisbon, Portugal, 4–8 Sept 2005, pp 813–816

Hönig F (2007) DRIVAWORK – driving under varying workload. A multi-modal stress database in the automotive context. Vortrag: HUMAINE plenary meeting, Paris, 6 Jun 2007

Ioannou S, Raouzaiou A, Tzouvaras V, Mailis T, Karpouzis K, Kollias S (2005) Emotion recognition through facial expression analysis based on a neurofuzzy network. Neural Netw (Elsevier) 18:423–435

Scherer KR, Ellgring H (2007) Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? Emotion 7(1):113–130

Sneddon I, McRorie M (2006) Perception of emotional expression. In: Poster presented at the 13th European conference on personality, Athens

Zara A, Maffiolo V, Martin J-C, Devillers L (2007) Collection and annotation of a corpus of human–human multimodal interactions: emotion and other anthropomorphic characteristics. In: Paiva A, Prada R, Picard R (eds) Affective computing and intelligent interaction, Lisbon, Sept 2007. Springer LNCS , Berlin, pp 464–475