Natural Affect Data - Collection & Annotation in a Learning Context

Shazia Afzal
Univ. of Cambridge
Computer Laboratory
Shazia.Afzal@cl.cam.ac.uk

Peter Robinson
Univ. of Cambridge
Computer Laboratory
Peter.Robinson@cl.cam.ac.uk

Abstract

Automatic inference of affect relies on representative data. For viable applications of such technology the use of naturalistic over posed data has been increasingly emphasised. Creating a repository of naturalistic data is however a massively challenging task. We report results from a data collection exercise in one of the most significant application areas of affective computing, namely computer-based learning environments. The conceptual and methodological issues encountered during the process are discussed, and problems with labelling and annotation are identified. A comparison of the compiled database with some standard databases is also presented.

1. Introduction

As emotion research gradually integrates with HCI studies and matures in application from mere prevention of usability problems to promoting richer user experiences, the need to capture 'pervasive emotion' [7] and also its context of occurrence is becoming an increasing concern. Existing databases are often oriented to prototypical representations of a few basic emotional expressions, being mostly posed or recorded in scripted situations. Such extreme expressions of affect rarely occur, if at all, in HCI contexts. The applicability of such data therefore becomes severely limited because of observed deviation from real-life situations [5, 7] and for our purpose their relevance to a learning situation like one-on-one interaction with a computer tutor. For developing applications that are generalisable to real-life situations there is now an increasing shift from easier to obtain posed data to more realistic naturally occurring data in the target scenarios. Dealing with the complexity associated with naturalistic data is however a significant problem. Nonverbal behaviour is rich, ambiguous and hard to validate, making naturalistic data collection and labelling a tedious, expensive and time-consuming exercise. In addition, lack of a consistent model of affect makes the abstraction of observed behaviour into appropriate labelling constructs very arbitrary.

This paper reports results from our data collection exercise in a potential application area – computer-based learning. The conceptual and methodological issues encountered during data collection are discussed, and problems with labelling and annotation are identified. Section 2 gives details about the data collection setup and experiment while Section 3 describes the three approaches to annotation on the compiled data. A comparison of our database against some standard databases is outlined in Section 4.

2. Data Collection

Our research involves modelling the affective aspects of learner experience in computer-based learning scenarios. As such we are interested in studying how non-verbal behaviour from multiple-cues like facial expressions, eye-gaze and head posture can be used to infer a learner's affective state during interaction and learning with a computer tutor. The final objective is to abstract this behaviour in terms of features that can enable automatic prediction and reliable computational modelling of affect states. The need for representative data is therefore essential in order to carry out realistic analyses, to develop appropriate techniques and eventually to perform validation of the inferences.

There is evidence that naturalistic head and facial expressions of affect differ in configuration and dynamics from posed/acted ones [6, 18]. These are in fact mediated by separate neural pathways and Ekman [10] identifies at least six characteristics that distinguish spontaneous from posed facial actions: morphology, symmetry, duration, speed of onset, coordination of apexes and ballistic trajectory. Moreover, there is an increasing emphasis on the role of situational context in the nature and meaning of emotion [19]. Ideally then, a database should depict naturalism, limited or no experimental control, and be contextually relevant. Since existing databases mostly include deliberately expressed emotions and are recorded in contexts that differ from their eventual application, their relevance to a naturalistic situation like learning with a computer is debatable conceptually, and as found practically [5, 7, 10]. To ensure ecological validity of our research, it was therefore necessary to study the affect patterns *in situ,* as they occur, for a more meaningful interpretation. This motivated our data collection exercise details of which are presented in the following sections.

978-1-4244-4799-2/09/$25.00 ©2009 IEEE

Figure 1: Examples of some facial expressions from our database collected in a learning scenario

2.1. Subjects

A total of eight subjects, three males and five females in the age group of 21 to 32, were recruited to serve as *encoders* of emotional behaviour. We will use the term encoders to denote these subjects in order to signify them as the source or examples of affective data obtained [10]. All were regular and proficient computer users (μ=20 hrs of computer usage per week). Two of the subjects wore glasses while one sported a beard. All subjects recorded as being happy, relaxed or in anticipation at the onset. They were informed that they would be video recorded during the interaction but remained naïve to the actual purpose of the experiment until after the experiment finished.

2.2. Setup

The recording setup was based on guidelines in Frank et al. [13]. The experiment was conducted in our usability lab which has a mock living room or personal office environment effect. It was chosen to facilitate video recording without compromising the naturalism of the desired behaviour. Standard computing equipment i.e. a desktop computer with a mouse and keyboard was used for the experiment. A video camera was mounted on top of the computer screen to allow video recording of the subjects' upper body especially the face.

2.3. Procedure

Subjects were run individually in the usability lab and were observed by the experimenter via a one-way mirror during the process. Formal consent for recording was taken in written from all subjects prior to the experiment. Subjects were video recorded while doing two tasks: an interactive map-based geography tutorial and a card matching activity. The tutorial enabled participants to read about countries and their locations followed by a test of their learning. There was no time limit on this task - participants took on average about 20 minutes to complete this activity. The second task was an adaptation of a card sorting activity meant to demonstrate the effect of situational anxiety on higher mental activities [21]. Cards having one, two, three or four of either squares, circles, crosses or triangles in red, green, blue or yellow were used - all figures on a card being alike and of same colour. Participants had to sort the cards against four category cards based on a changing criterion – like colour, shape, number or all in order. This task triggers a cycle of anxiety by inhibiting reflective intelligence leading to lowered performance and thus, decreased motivation [21]. We wanted to observe if this cumulative effect of anxiety on task performance was accompanied by any observable patterns in facial expressions.

The two tasks were chosen to ensure a variety of emotion expressions and were sequentially varied across subjects. The card activity contained three triggers/events: the screen blanking out for five seconds, a match not being possible and variation in feedback/scoring. These are not dramatic deviations from the task but rather represent some common interaction events and were presented only once per participant during the game interaction (order-varied). The session finished with a semi-structured interview and subsequently, self-annotation of videos (as discussed in Section 3.1).

2.4. Discussion

Approximately four hours of video data was collected from the eight subjects. Overall there was significant variability in the emotional behaviour of subjects. Individual differences in expressivity were striking. Some subjects were animated and displayed a wide range of expressions while others were notably inexpressive. There was difference even in the way subjects reacted to the triggered events. Manifested behaviour was found to be strongly related to dispositional traits of emotional expressivity and attitude. We also observed that consistently across our encoder (subject) group, more emotional expressions occurred during the card game than during the tutorial. This implies that the task difference has a substantial impact on the nonverbal behaviour of individuals.

As individual expressivity and task differences appeared as the two major factors affecting emotional behaviour, it is reasonable to suggest that emotion inference technology will need to address these in both design and function. An in-depth discussion of these issues based on a qualitative analysis of the recorded data appears in [1].

3. Annotation & Labelling

Our motivation to collect this data was to get an idea of the range of emotional behaviour that occurs in a learning scenario. The long term objective is to make use of affective cues to automatically predict a learners' emotional state. Automatic prediction using machine learning relies on extensive training data which in this case implies preparation of labelled representative data. This requires observational assessments on data to be represented in a quantifiable manner via annotation. It involves developing a protocol to catalogue observations and to represent the behaviour of interest using an appropriate coding scheme in terms of desired labelling constructs. Our annotation method evolved from various domain relevant decisions related to the choice of labelling constructs and modality, anticipated technical constraints in target scenario, relation to context and ease of interpretation. Before elaborating on the annotation process itself, we outline the choices and practices from nonverbal behaviour research that provide the framework for our data annotation.

Coding scheme

Coding schemes can range along a continuum, with one end anchored by physiologically based schemes, and the other end by socially based coding schemes [3]. Examples of physiologically based coding schemes are FACS, MAX and MPEG-4 which even though more objective and comprehensive, are complex, require extensive training and involve specialised procedures. Since our goal is to examine emotional behaviour we adopt a socially based scheme which is defined as an observational system that examines behaviours or messages that have more to do with social categories of interaction like smiling rather than with physiological elements of behaviour like amplitude [22].

Level of Measurement

This refers to the level of abstraction adopted while measuring behaviour and can be at a macro level to capture perceptual judgements and social meaning of behaviour, or at a micro level for more concrete assessments [22]. We reconcile the two frames of references in our approach by following hierarchical labelling on our data. An inferential level annotation by extracting emotionally salient segments is followed by two levels of a more focused coding on these.

Coding Unit

The coding unit refers to the decisions about when to code within an interaction and the length of time the observation should last [3]. It has two broad variants - event based and interval based. Choosing one over the other depends upon the research view and the level of accuracy required, complexity of the coding scheme and the frequency of behaviour occurrence [3]. We used interval-based coding to allow a systematic and consistent observation in our first annotation round, but given the nature of the data, as discussed in Section 3.1, were forced to adopt the event based coding eventually.

Labelling Construct

Annotation schemes commonly employ either categorical, dimensional or appraisal based labelling approaches [7]. In addition, free-response labelling may also be used for richer descriptions. We use a variant of categorical labelling where coders are asked to choose from pre-selected domain relevant emotional descriptors namely confused, interested, surprised, happy, bored, and annoyed. These descriptors refer to non-basic affective-cognitive states and are pertinent in learning situations. Coders are given the alternative to define their own category or label under a residual 'Other' option if none of the provided labels reflect their judgement. This ensures a degree of flexibility in coding. An emotion list is provided at the onset of coding for familiarisation with everyday emotion terms.

Coders

Selecting coders is an important aspect of designing annotation studies as they should be able to discern meaning from behaviour and make judgements effectively. We have attempted three modes of annotation with respect to coders: self-annotation by encoders themselves, experts and non-experts.

Reliability Measures

Inter-coder reliability measures for nominal data include raw-agreement, Scott's pi, Cohen's kappa, Fleiss' kappa, and Krippendorff's alpha [14]. Our approach involves multiple-coders rating multiple categories - often uneven across coders, and as such we use Fleiss' kappa to report inter-coder reliability. Fleiss' kappa is a statistical measure that calculates the degree of agreement in classification over that expected by chance and is scored as a number between 0 and 1, where 1 indicates perfect agreement [12].

Having set the scope of our annotation in terms of general methodological decisions, we now describe the three iterations of annotation that our data underwent.

3.1. First Annotation

Design

Self-annotation can serve as a triangulation method while analysing coding results by allowing comparisons between felt emotions and observed behaviour. Given our specific research setup and the type of labelled data we sought, none of the standard self-report instruments were found suitable [15]. Self-annotation was therefore implemented using an interval-based coding system where subjects were prompted to rate each of the selected emotion categories after every 20 seconds of elapsed video while also allowing a free-response option to allow subjective descriptions. Annotation was implemented to allow a split-screen viewing of recorded behaviour with the time synchronised interaction record obtained via screen capture to encourage context-sensitive judgment.

Results

The purpose of obtaining self-report was to get a subjective account of emotional behaviour. Observation of the labelling process however, indicated otherwise. Although participants responded differently to watching their own expressions - some surprised, mimicking and laughing at themselves, and others embarrassed, rushing through the video; the reactions did not suggest that they associated a subjective feeling but rather interpreted the expression as they might if it belonged to another in a social setting. This level of cognitive mediation was perceived as confounding the self-labelling purpose. It seemed that subjects were more interested in 'watching' themselves and rushed through the coding part. They also complained that 20 seconds was a very tiny interval and that 'nothing major' was happening. Three subjects left the coding mid-way complaining of boredom. For these reasons, the self-annotation was considered unreliable and was discarded. It is possible that our choice of self-report assessment was inefficient and a better strategy could be devised. However two important observations that resulted from this attempt shaped the next level of annotation:

1) Emotional behaviour in the videos was subtle and gradual making interval-based coding extremely tedious. Switching to event-based coding was deemed appropriate for maximising the value of annotation.

2) To improve reliability of annotation, a more objective labelling using multiple external coders was adopted.

3.2. Second Annotation

Design

The original videos were segmented into 105 non-neutral segments using a video analysis tool ELAN[1]. The mean duration of extracted clips was 101 frames (σ = 73.63), ranging from a min. of 17 frames to a max. of 480 frames. The segmentation was based on changes in the blanket expression, where behaviour seemed consistent over a period of time. This essentially meant extracting portions of video that contained emotional behaviour. Care was taken to preserve the temporal boundaries while demarcating segments. At 30 fps, this reduced the original video corpus of approximately 4 hours to less than 6 minutes. This was a substantial gain in required annotation effort for subsequent labelling.

Results

Three expert coders, 2 male 1 female, labelled the 105 pre-segmented clips independently. Coders could replay a video as many times as they wished. A primary and optional secondary emotion label was allowed for each video clip. Enforcing a simple majority rule resulted in 75% of videos getting classified into one of our pre-selected emotion categories. Table 1 (column A) summarises the distribution of emotion categories obtained this way when at least two out of the three coders agreed.

Taking primary labels into account, Fleiss' overall kappa was 0.35 indicating fair agreement - agreement by chance was ruled out, but weakly. Given the low inter-coder reliability, the labelling results remained questionable. Moreover, the expert coders indicated that the video clips often displayed multiple emotions and that a second level of segmentation would improve judgement accuracy. A finer level of further segmentation increased the total number of video clips from 105 to 247. A third level annotation on these was designed, as outlined in the next section.

Table 1: Distribution of emotions categories

Annotation(s)	A		B	
	3 experts, 105 clips		108 coders, 247 clips	
	No.	***% age***	***No.***	***% age***
Confused	26	24.8 %	73	29.6 %
Interested	18	17.1 %	35	14.2 %
Surprised	12	11.4 %	40	16.2 %
Bored	5	4.8 %	19	7.7 %
Happy	16	15.2 %	35	14.2 %
Annoyed	0	0 %	13	5.3 %
Neutral	3	2.9 %	29	11.7 %
Other	25	23.9 %	3	1.2 %

3.3. Third Annotation

Design

The corpus now consisted of 247 video clips with a mean duration of 84.34 frames (σ = 55.80), ranging from a minimum of 12 frames to maximum of 479 frames. An online interface was set-up to facilitate access to a large number of coders. The coding scheme was modified so that for each video clip coders were required to mark the following: the emotion they attributed to the video, their confidence level (from 1-10) and whether they could perceive more than one emotion in the clip. The decision time for emotion judgement was also recorded. A video clip was played only once in order to get the initial reaction and also to control effects of replaying across coders. The focus at this level of annotation was to analyse emotion judgements from a large number of coders and improve annotation results.

Results

108 coders, 39 male 69 female, signed up for the online study and coded an average 20 videos each. They were aged between 18 and 56 (μ = 28.28, σ = 6.20) and were of diverse ethnicities and background. A total of 2221 annotations was obtained so that each video was coded on average 8.99 times (σ = 0.13).

Emotion labels present under 'Other' category were parsed using emotion taxonomies, GALC [20] and Mind Reading [4] in order to group semantically similar terms into macro-classes. For example, pleased, amused, and enjoying, were grouped together under 'Happy'.

Inter-coder reliability estimated using Fleiss' weighted kappa for multiple ratings per video with

[1] http://www.lat-mpi.eu/tools/tools/elan

multiple raters [12] was 0.20 overall, indicating slight agreement. Individual kappa agreements for the emotion categories are listed in Table 2.

Table 2: Fleiss' kappa for each of the emotion categories

Confused	0.18	Happy	0.52
Interested	0.09	Annoyed	0.10
Surprised	0.20	Neutral	0.17
Bored	0.13	Other	0.05

Only 'Happy' shows a good agreement while the others show marginal kappa values. In fact, if we look at the decision time across emotion categories in Figure 2 below, it appears that videos classified as Happy are quicker to recognise than others. Videos classified as 'Other' show relatively longer decision times.

As the reliability results did not meet a fair agreement criterion, we adopted a weighted system for classification of videos into emotion categories using the coders' confidence ratings. Emotion labels were

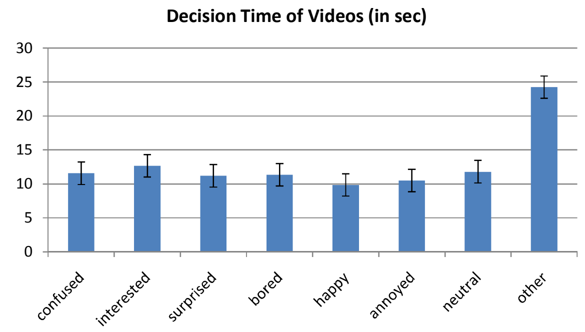

Figure 2: Emotion wise decision times (in seconds)

assigned a weight equivalent to a coders' confidence level. The maximum weighted emotion label was then taken as the true label for a video clip. For example, if a video clip A was coded as 'happy' with confidence 9 by Coder1, 'happy' with confidence 7 by Coder2, 'confused' with confidence 1 and say 'surprised' with confidence 9; the clip would be classified as 'happy' since the total confidence weight for 'happy' is higher (9+7) than each of the others. The distribution of video clips across emotion categories following this is shown in Table 1 (column B). Figure 3 shows the average duration of the videos across the different emotion

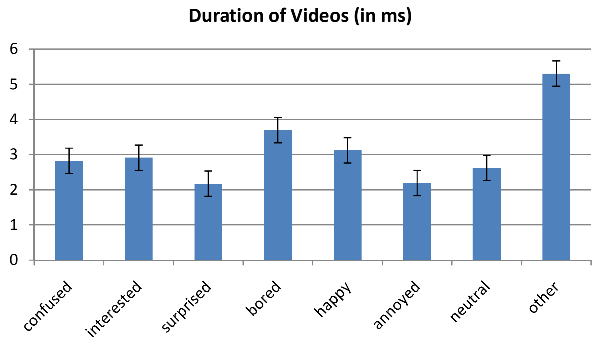

Figure 3: Emotion wise duration of videos (in ms)

categories.

3.4. Discussion

Having clearly labelled samples is a pre-requisite for designing automatic classifiers. Our experiment suggests that this is indeed very hard to obtain from naturalistic data. Even for a human expert, it is difficult to define what constitutes an emotion. Segmenting the original videos into emotionally salient clips was the most labour-intensive and time-consuming process. Demarcating the beginning and end of emotional expressions is incredibly hard as they often overlap, co-occur or blend subtly into a background expression. Pre-segmentation of videos should ideally be validated by a second and if possible, more coders.

Re-visiting the data often changes judgements as familiarity habituates a coder to the range of facial signs in encoders. The whole process is unavoidably subjective and therefore dependent on the affect decoding skills and experience of coders. Table 4 for example shows gender-wise results for confidence ratings, decision time in making emotion judgements, and marking of more than one emotion. Female coders appear to be more confident, arrive at judgements faster but perceive more than one emotion consistently.

Coders, expert as well non-expert, often used a combination of labels and even phrases to express their judgements. The 'Other' category was liberally used which reveals the dependence of coders' active vocabulary on annotation. This supports balancing free-form responses with fixed-choice alternatives to maximise accuracy while ensuring a degree of standardisation. Having taxonomies that allow mapping of free-form lexical emotion labels into different levels or groups of emotions would be of great help to standardise annotation results. Taxonomies like the GALC [20] and Mind Reading [4] though not entirely comprehensive as yet, are good examples.

Table 3: Gender-wise coding results

Confidence		**Decision Time**		**>1 emotion**	
m	*f*	*m*	*f*	*m*	*f*
7.41 (1.89)	7.57 (1.89)	12.47 (13.67)	11.52 (8.33)	36.4%	63.6%

Ambiguity in emotion judgements is another factor that comes to the fore. 38.7% of the total videos were perceived as containing more than one emotion. Female coders on average made higher use of this option than males.

Multiple layers of annotation may help to reduce the subjectivity of annotations. Abrilian et al's [2] multi-level annotation framework is exemplary in that it combines emotion, context and multimodal annotations to overcome issues related to temporality and abstraction. Coding-time and cost however remain the main constraints as in any comprehensive coding technique.

Additionally, low inter-coder reliabilities, rather than being an error of measure as one could interpret, are in fact an acknowledged observation reported for naturalistic data [2, 7]. What is important is to understand how we can decide on an optimal metric of recognition accuracy for evaluating automatic classifiers, when we lack a reliable and objective ground-truth in the first place.

4. Related Work

Space constraints presenting a detailed comparison against all published visual databases here. Instead, we present a comparison of our database with some widely used ones in Table 4. Our database is one of the first published naturalistic databases obtained in the target application scenario. It differs from works in the ITS community [c.f. 8] by the intentional absence of 'intelligent adaptive tutoring'. This was done in order to study the affective behaviour during self-regulated non-adaptive learning with computers and to reduce the complexity in interpreting it by limiting the influence of additional factors that may arise in an adaptive interaction.

5. Summary & Conclusions

Emotion and expressivity have contextual significance so that if we adopt an application-oriented view, reliance on re-usable general databases is perhaps of limited value. For affect recognition technology to reliably operate in target applications, we need context-specific corpora to serve not only as repositories of sample data but importantly to shape our understanding of the problem itself. This paper has described one such attempt to capture naturalistic emotional data in a computer based learning scenario. We have discussed our data collection and annotation procedures in detail and have discussed initial observations.

6. Acknowledgements

We would like to thank David Good and Alan Blackwell for stimulating discussions, Ian Davies for technical help, and all our participants and coders. This research is supported by the Gates Cambridge Trust.

References

[1] Afzal, S., Morrison, C. & Robinson, P. (2009). Intentional affect: An alternative notion of affective interaction with a machine, HCI '09, Cambridge, UK

[2] Abrilian, S., Devillers, L., Buisine, S., Martin, J-C (2005). "EmoTV1: Annotation of Real-life Emotions for the Specification of Multimodal Affective Interfaces", HCI International.

[3] Bakeman R., Gothman J.M. (1997). *Observing interaction: An introduction to sequential analysis*, Cambridge University Press, UK.

[4] Baron-Cohen, S., Golan O., Wheelwright S., & Hill J. (2004). Mind Reading: The Interactive Guide to Emotions, Jessica Kingsley Publishers, London.

[5] Batliner A., Fischer K., Huber R., Spilker J., Noth E. (2003). How to Find Trouble in Communication, Speech Communication, 40 (1-2), pp. 117-143

[6] Cohn J. F. & Schmidt, K. L. (2004). The timing of facial motion in posed and spontaneous smiles. Wavelets, Multiresolution and Information Processing, 2, pp. 1-12.

[7] Cowie, R., Douglas-Cowie E. & Cox C. (2005). Beyond emotion archetypes: Databases for emotion modelling using neural networks, Neural Networks, 18, 371-388

[8] D'Mello S.K., Craig S.D., Gholson B., Franklin S., Picard R.W. & Graesser A.C. (2005). Integrating Affect Sensors in an Intelligent Tutoring System. Workshop on Affective Interactions: The Computer in the Affective Loop Workshop, IUI '05.

[9] Douglas-Cowie, E. et al. (2004). D5c Preliminary plans for exemplars: Databases, HUMAINE deliverable.

[10] Ekman, P. & Rosenberg E.L. (1997). *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using Facial Action Coding System (FACS)*, Oxford University Press.

[11] El Kaliouby, R. & Teeters, A. (2007). Eliciting, Capturing and Tagging Spontaneous Facial Affect in Autism Spectrum Disorder, ICMI'07, Aichi, Japan.

[12] Fleiss, J.L., Levin, B. & Paik, M.C. (2003). Statistical Methods for Rates & Proportions, 3rd Edition, Wiley, NJ.

[13] Frank, M.G., Juslin, P.N. & Harrigan, J.A. (2005). Technical Issues in Recording Nonverbal Behaviour, In Harrigan, J.A. (Ed.). *The New Handbook of Methods in Nonverbal Behaviour Research*, Oxford University Press

[14] Hayes, A.F. & Krippendorff, K. (2007). Answering the Call for a Standard Reliability Measure for Coding Data, Communication Methods and Measures, 1(1), 77–89

[15] Isomursu M., Tahti M., Vainamo S., & Kuutti K. (2007). Experimental evaluation of five methods for collecting emotions in field settings with mobile application, Journal of Human-Computer Studies, 65, pp. 404-418.

[16] Kanade, T., Cohn, J. & Tian, Y.L. (2000). Comprehensive database for facial expression analysis, In Proc. IEEE FG '00.

[17] Pantic, M., Valstar, M., Rademaker, R., & Maat, L. (2005). Web-based database for facial expression analysis. In Proc. IEEE ICME'05.

[18] Pantic M. & Patras I. (2006). Dynamics of Facial Expression: Recognition of Facial Actions and Their Temporal Segments from Face Profile Image Sequences, IEEE Trans. In Systems, Man, and Cybernetics, 36(2).

[19] Russell J.A., Fernandez-Dols J.M. (1997). *The psychology of facial expression.* Cambridge, MA: Cambridge UP.

[20] Scherer, K. L. (2005). What are emotions? And how can they be measured? Social Science Information, 44(4).

[21] Skemp RR. (1971). *The Psychology of Learning Mathematics*, Hillsdale, NJ: Erlbaum

[22] White, C.H. & Sargent, J. (2005). Researcher Choices and Practices in the Study of Nonverbal Communication, In Manusov V.L (Ed.). *Sourcebook of Nonverbal Measures: Going Beyond Words*. Lawrence Erlbaum Associates.

Table 4: Comparison of our compiled database (CAL) with some standard visual databases

Properties	**Cohn-Kanade**	**MMI**	**Mind Reading DVD**	**Belfast Naturalistic**	**MIT-Groden-Autism**	**CAL**
General						
Availability	Public / Licensed	Public / Licensed	Nominal fee	Public / Licensed	Protected for privacy reasons	Protected for privacy reasons
Emotion labels	No	Yes	Yes	Yes	Yes	Yes
FACS-coded	Yes	Yes	No	No	No	No
Format	Downloadable	Web-based, Downloadable	DVD	CDs	-	CDs
Author(s)	Kanade et al. [16]	Pantic et al. [17]	Baron-Cohen et al. [4]	Douglas-Cowie, E. et al.[7, 9]	El Kaliouby & Teeters [11]	
Elicitation Method						
Spontaneity	Posed	Posed	Posed	Naturalistic	Induced	Naturalistic
Expt. Control	Directed	Directed	Unconstrained	Unconstrained	Unconstrained	Unconstrained
Scenario	Instructed by expert	Instructed by expert	Example scenarios	Sensitive Artificial Listener (SAL), Interviews	Games & interaction scenarios	Computer-assisted learning environment
Context	Individual	Individual	Individual	Social Interactive	Social (Dyadic)	Human-Computer Interaction
Emotional Content						
Modalities	Visual	Visual	Visual	Audio-Visual	Visual	Visual
No. of videos	2105	848	2742	239 sequences	2090	4 hrs original; 247 clips
Min-Max duration	0.3 - 2.0 sec	1.66 – 21.6 sec	5.0 - 8.0 sec	10 – 60 sec	? - 10.9 sec	0.4 - 15.9 sec
Resolution	640 x 480	720 x 576	320 x 240	-	-	320 x 240
Frame Rate	-	24 fps	30 fps	-	-	30fps
Lighting	Uniform	Uniform	Uniform	Variable	Variable	Variable
Pose	Frontal	Frontal + Profile View	Frontal	Frontal	Frontal	Frontal
Initial frame	Neutral	Neutral	Non-neutral	Non-neutral	Non-neutral	Non-neutral
Rigid head motion	No	No	Yes	Yes	Yes	Yes
Occlusion	No	No	No	Yes	Yes	Yes
Talking	No	No	No	Yes	Yes	Yes
Encoders						
No. of subjects	210	19	30	125	8	8
Gender (M : F)	31 : 69	10 : 9	15 : 15	31 : 94	1 : 7	3 : 5
Age-group	18-50 yrs	19-62 yrs	16-60 yrs	-	18-20 yrs	21-32 yrs
Ethnicity	Diverse	Diverse	Diverse	-	-	Diverse
Glasses	No	Yes	No	-	Yes	Yes
Facial Hair	No	Yes	No	-	-	Yes
Labelling						
Coding Model	FACS	FACS	affect labels	affect labels; dimensions	affect labels	affect labels
No. of coders	2 FACS experts	2 FACS experts	10	6	10 (pre-segmentation by an expert)	108 (pre-segmentation by an expert)
Inter-coder reliability	0.86 Cohen's kappa	consensus	8/10 raw agreement	-	8/10 raw agreement	0.20 Fleiss' kappa
Emotional Content	Six basic emotions: Joy, surprise, anger, fear, disgust, sad	Six basic emotions: Joy, surprise, anger, fear, disgust, sad	412 affective-cognitive states	48 Emotion words, valence, activation, intensity	Agreeing, disagreeing, interested, thinking, confused, concerned, happy, sad, surprise, anger	Affective-cognitive states / Non-basic emotions: Confused, interested, surprised, bored, happy, annoyed, other
Multiple-labels	NA	NA	No	Yes	Yes	Yes
Context info	No	No	No	Yes	-	Yes